

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
24 October 2002 (24.10.2002)

PCT

(10) International Publication Number  
**WO 02/083709 A2**

(51) International Patent Classification<sup>7</sup>: **C07K**

(21) International Application Number: PCT/US02/10818

(22) International Filing Date: 5 April 2002 (05.04.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

|            |                            |    |
|------------|----------------------------|----|
| 60/282,119 | 6 April 2001 (06.04.2001)  | US |
| 60/283,588 | 13 April 2001 (13.04.2001) | US |
| 60/283,759 | 13 April 2001 (13.04.2001) | US |
| 60/285,589 | 20 April 2001 (20.04.2001) | US |
| 60/287,037 | 27 April 2001 (27.04.2001) | US |
| 60/287,036 | 27 April 2001 (27.04.2001) | US |
| 60/288,608 | 4 May 2001 (04.05.2001)    | US |
| 60/288,712 | 4 May 2001 (04.05.2001)    | US |
| 60/289,909 | 9 May 2001 (09.05.2001)    | US |
| 60/292,246 | 17 May 2001 (17.05.2001)   | US |

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): RECIPON, Shirley, A. [US/US]; 85 Fortuna Avenue, San Francisco, CA 94115 (US). BURRILL, John, D. [US/US]; 2218 Brewster Avenue, Redwood City, CA 94062 (US). MARCUS, Gregory, A. [US/US]; 73 Maple Way, San Carlos, CA 94070 (US). ZINGLER, Kurt, A. [US/US]; 723 Ashbury Street #C, San Francisco, CA 94117 (US). TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). THORNTON, Michael [US/US]; 9 Medway Road, Woodside, CA 94062 (US). BOROWSKY, Mark, L. [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). BURFORD, Neil [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). LEE, Soo, Yeun [KR/US]; 40 Westdale Avenue, Daly City, CA 94015 (US). BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). HAFALIA, April, J., A. [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). YAO, Monique, G. [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). RAMKUMAR, Jayalaxmi [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). WALIA, Narinder, K. [US/US]; 890

Davis Street A # 205, San Leandro, CA 94577 (US). LU, Dyung, Aina, M. [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). ARVIZU, Chandra, S. [US/US]; 1706 Morocco Drive, San Jose, CA 95125 (US). ISON, Craig, H. [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US). DING, Li [CN/US]; 3353 Alma Street # 146, Palo Alto, CA 94306 (US). LU, Yan [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). GURURAJAN, Rajagopal [IN/US]; 5591 Dent Avenue, San Jose, CA 95118 (US). WALSH, Roderick, T. [IE/GB]; 8 Boundary Court St. Lawrence Road, Canterbury Kent, Antrim CT1 3EZ (GB). GANDHI, Ameena, R. [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). SWARNAKAR, Anita [CA/US]; 8 Locksley Avenue #5D, San Francisco, CA 94122 (US). FORSYTHE, Ian, J. [US/US]; 308 Roble Avenue, Redwood City, CA 94061 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). AU-YOUNG, Janice, K. [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). ELLIOTT, Vicki, S. [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). LEE, Sally [US/US]; 3643 26th Street, San Francisco, CA 94110 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: KINASES AND PHOSPHATASES

(57) Abstract: The invention provides human kinases and phosphatases (KPP) and polynucleotides which identify and encode KPP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of KPP.

WO 02/083709 A2

## KINASES AND PHOSPHATASES

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of kinases and phosphatases  
5 and to the use of these sequences in the diagnosis, treatment, and prevention of cardiovascular  
diseases, immune system disorders, neurological disorders, disorders affecting growth and  
development, lipid disorders, cell proliferative disorders, and cancers, and in the assessment of the  
effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of  
kinases and phosphatases.

10

### BACKGROUND OF THE INVENTION

Reversible protein phosphorylation is the ubiquitous strategy used to control many of the  
intracellular events in eukaryotic cells. It is estimated that more than ten percent of proteins active in  
a typical mammalian cell are phosphorylated. Kinases catalyze the transfer of high-energy phosphate  
15 groups from adenosine triphosphate (ATP) to target proteins on the hydroxyamino acid residues  
serine, threonine, or tyrosine. Phosphatases, in contrast, remove these phosphate groups.  
Extracellular signals including hormones, neurotransmitters, and growth and differentiation factors can  
activate kinases, which can occur as cell surface receptors or as the activator of the final effector  
protein, as well as other locations along the signal transduction pathway. Cascades of kinases occur,  
20 as well as kinases sensitive to second messenger molecules. This system allows for the amplification  
of weak signals (low abundance growth factor molecules, for example), as well as the synthesis of  
many weak signals into an all-or-nothing response. Phosphatases, then, are essential in determining  
the extent of phosphorylation in the cell and, together with kinases, regulate key cellular processes  
such as metabolic enzyme activity, proliferation, cell growth and differentiation, cell adhesion, and cell  
25 cycle progression.

### KINASES

Kinases comprise the largest known enzyme superfamily and vary widely in their target  
molecules. Kinases catalyze the transfer of high energy phosphate groups from a phosphate donor to  
a phosphate acceptor. Nucleotides usually serve as the phosphate donor in these reactions, with most  
30 kinases utilizing adenosine triphosphate (ATP). The phosphate acceptor can be any of a variety of  
molecules, including nucleosides, nucleotides, lipids, carbohydrates, and proteins. Proteins are  
phosphorylated on hydroxyamino acids. Addition of a phosphate group alters the local charge on the  
acceptor molecule, causing internal conformational changes and potentially influencing intermolecular

contacts. Reversible protein phosphorylation is the primary method for regulating protein activity in eukaryotic cells. In general, proteins are activated by phosphorylation in response to extracellular signals such as hormones, neurotransmitters, and growth and differentiation factors. The activated proteins initiate the cell's intracellular response by way of intracellular signaling pathways and second messenger molecules such as cyclic nucleotides, calcium-calmodulin, inositol, and various mitogens, that regulate protein phosphorylation.

Kinases are involved in all aspects of a cell's function, from basic metabolic processes, such as glycolysis, to cell-cycle regulation, differentiation, and communication with the extracellular environment through signal transduction cascades. Inappropriate phosphorylation of proteins in cells has been linked to changes in cell cycle progression and cell differentiation. Changes in the cell cycle have been linked to induction of apoptosis or cancer. Changes in cell differentiation have been linked to diseases and disorders of the reproductive system, immune system, and skeletal muscle.

There are two classes of protein kinases. One class, protein tyrosine kinases (PTKs), phosphorylates tyrosine residues, and the other class, protein serine/threonine kinases (STKs), phosphorylates serine and threonine residues. Some PTKs and STKs possess structural characteristics of both families and have dual specificity for both tyrosine and serine/threonine residues. Almost all kinases contain a conserved 250-300 amino acid catalytic domain containing specific residues and sequence motifs characteristic of the kinase family. The protein kinase catalytic domain can be further divided into 11 subdomains. N-terminal subdomains I-IV fold into a two-lobed structure which binds and orients the ATP donor molecule, and subdomain V spans the two lobes. C-terminal subdomains VI-XI bind the protein substrate and transfer the gamma phosphate from ATP to the hydroxyl group of a tyrosine, serine, or threonine residue. Each of the 11 subdomains contains specific catalytic residues or amino acid motifs characteristic of that subdomain. For example, subdomain I contains an 8-amino acid glycine-rich ATP binding consensus motif, subdomain II contains a critical lysine residue required for maximal catalytic activity, and subdomains VI through IX comprise the highly conserved catalytic core. PTKs and STKs also contain distinct sequence motifs in subdomains VI and VIII which may confer hydroxyamino acid specificity.

In addition, kinases may also be classified by additional amino acid sequences, generally between 5 and 100 residues, which either flank or occur within the kinase domain. These additional amino acid sequences regulate kinase activity and determine substrate specificity. (Reviewed in Hardie, G. and S. Hanks (1995) The Protein Kinase Facts Book, Vol I, pp. 17-20 Academic Press, San Diego CA.). In particular, two protein kinase signature sequences have been identified in the kinase domain, the first containing an active site lysine residue involved in ATP binding, and the

second containing an aspartate residue important for catalytic activity. If a protein analyzed includes the two protein kinase signatures, the probability of that protein being a protein kinase is close to 100% (PROSITE: PDOC00100, November 1995).

### Protein Tyrosine Kinases

5 Protein tyrosine kinases (PTKs) may be classified as either transmembrane, receptor PTKs or nontransmembrane, nonreceptor PTK proteins. Transmembrane tyrosine kinases function as receptors for most growth factors. Growth factors bind to the receptor tyrosine kinase (RTK), which causes the receptor to phosphorylate itself (autophosphorylation) and specific intracellular second messenger proteins. Growth factors (GF) that associate with receptor PTKs include epidermal GF,  
10 platelet-derived GF, fibroblast GF, hepatocyte GF, insulin and insulin-like GFs, nerve GF, vascular endothelial GF, and macrophage colony stimulating factor.

Nontransmembrane, nonreceptor PTKs lack transmembrane regions and, instead, form signaling complexes with the cytosolic domains of plasma membrane receptors. Receptors that function through non-receptor PTKs include those for cytokines and hormones (growth hormone and  
15 prolactin), and antigen-specific receptors on T and B lymphocytes.

Many PTKs were first identified as oncogene products in cancer cells in which PTK activation was no longer subject to normal cellular controls. In fact, about one third of the known oncogenes encode PTKs. Furthermore, cellular transformation (oncogenesis) is often accompanied by increased tyrosine phosphorylation activity (Charbonneau, H. and N.K. Tonks (1992) Annu. Rev.  
20 Cell Biol. 8:463-493). Regulation of PTK activity may therefore be an important strategy in controlling some types of cancer.

### Protein Serine/Threonine Kinases

Protein serine/threonine kinases (STKs) are nontransmembrane proteins. A subclass of STKs are known as ERKs (extracellular signal regulated kinases) or MAPs (mitogen-activated  
25 protein kinases) and are activated after cell stimulation by a variety of hormones and growth factors. Cell stimulation induces a signaling cascade leading to phosphorylation of MEK (MAP/ERK kinase) which, in turn, activates ERK via serine and threonine phosphorylation. A varied number of proteins represent the downstream effectors for the active ERK and implicate it in the control of cell proliferation and differentiation, as well as regulation of the cytoskeleton. Activation of ERK is  
30 normally transient, and cells possess dual specificity phosphatases that are responsible for its down-regulation. Also, numerous studies have shown that elevated ERK activity is associated with some cancers. Other STKs include the second messenger dependent protein kinases such as the cyclic-AMP dependent protein kinases (PKA), calcium-calmodulin (CaM) dependent protein kinases,



and the mitogen-activated protein kinases (MAP); the cyclin-dependent protein kinases; checkpoint and cell cycle kinases; Numb-associated kinase (Nak); human Fused (hFu); proliferation-related kinases; 5'-AMP-activated protein kinases; and kinases involved in apoptosis.

One member of the ERK family of MAP kinases, ERK 7, is a novel 61-kDa protein that has  
5 motif similarities to ERK1 and ERK2, but is not activated by extracellular stimuli as are ERK1 and ERK2 nor by the common activators, c-Jun N-terminal kinase (JNK) and p38 kinase. ERK7 regulates its nuclear localization and inhibition of growth through its C-terminal tail, not through the kinase domain as is typical with other MAP kinases (Abe, M.K. (1999) *Mol. Cell. Biol.* 19:1301-1312).

10 The second messenger dependent protein kinases primarily mediate the effects of second messengers such as cyclic AMP (cAMP), cyclic GMP, inositol triphosphate, phosphatidylinositol, 3,4,5-triphosphate, cyclic ADP ribose, arachidonic acid, diacylglycerol and calcium-calmodulin. The PKAs are involved in mediating hormone-induced cellular responses and are activated by cAMP produced within the cell in response to hormone stimulation. cAMP is an intracellular mediator of  
15 hormone action in all animal cells that have been studied. Hormone-induced cellular responses include thyroid hormone secretion, cortisol secretion, progesterone secretion, glycogen breakdown, bone resorption, and regulation of heart rate and force of heart muscle contraction. PKA is found in all animal cells and is thought to account for the effects of cAMP in most of these cells. Altered PKA expression is implicated in a variety of disorders and diseases including cancer, thyroid disorders,  
20 diabetes, atherosclerosis, and cardiovascular disease (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, New York NY, pp. 416-431, 1887).

The casein kinase I (CKI) gene family is another subfamily of serine/threonine protein kinases. This continuously expanding group of kinases have been implicated in the regulation of numerous cytoplasmic and nuclear processes, including cell metabolism and DNA replication and  
25 repair. CKI enzymes are present in the membranes, nucleus, cytoplasm and cytoskeleton of eukaryotic cells, and on the mitotic spindles of mammalian cells (Fish, K.J. et al. (1995) *J. Biol. Chem.* 270:14875-14883).

The CKI family members all have a short amino-terminal domain of 9-76 amino acids, a highly conserved kinase domain of 284 amino acids, and a variable carboxyl-terminal domain that ranges  
30 from 24 to over 200 amino acids in length (Cegielska, A. et al. (1998) *J. Biol. Chem.* 273:1357-1364). The CKI family is comprised of highly related proteins, as seen by the identification of isoforms of casein kinase I from a variety of sources. There are at least five mammalian isoforms,  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ , and  $\epsilon$ . Fish et al. identified CKI-epsilon from a human placenta cDNA library. It is a basic protein of 416

amino acids and is closest to CKI-delta. Through recombinant expression, it was determined to phosphorylate known CKI substrates and was inhibited by the CKI-specific inhibitor CKI-7. The human gene for CKI-epsilon was able to rescue yeast with a slow-growth phenotype caused by deletion of the yeast CKI locus, HRR250 (Fish et al., *supra*).

- 5       The mammalian circadian mutation tau was found to be a semidominant autosomal allele of CKI-epsilon that markedly shortens period length of circadian rhythms in Syrian hamsters. The tau locus is encoded by casein kinase I-epsilon, which is also a homolog of the *Drosophila* circadian gene double-time. Studies of both the wildtype and tau mutant CKI-epsilon enzyme indicated that the mutant enzyme has a noticeable reduction in the maximum velocity and autophosphorylation state.
- 10   Further, *in vitro*, CKI-epsilon is able to interact with mammalian PERIOD proteins, while the mutant enzyme is deficient in its ability to phosphorylate PERIOD. Lowrey et al. have proposed that CKI-epsilon plays a major role in delaying the negative feedback signal within the transcription-translation-based autoregulatory loop that composes the core of the circadian mechanism. Therefore the CKI-epsilon enzyme is an ideal target for pharmaceutical compounds influencing circadian rhythms, jet-lag
- 15   and sleep, in addition to other physiologic and metabolic processes under circadian regulation (Lowrey, P.L. et al. (2000) *Science* 288:483-491).

- Homeodomain-interacting protein kinases (HIPKs) are serine/threonine kinases and novel members of the DYRK kinase subfamily (Hofmann, T.G. et al. (2000) *Biochimie* 82:1123-1127). HIPKs contain a conserved protein kinase domain separated from a domain that interacts with
- 20   homeoproteins. HIPKs are nuclear kinases, and HIPK2 is highly expressed in neuronal tissue (Kim, Y.H. et al. (1998) *J. Biol. Chem.* 273:25875-25879; Wang, Y. et al. (2001) *Biochim. Biophys. Acta* 1518:168-172). HIPKs act as corepressors for homeodomain transcription factors. This corepressor activity is seen in posttranslational modifications such as ubiquitination and phosphorylation, each of which are important in the regulation of cellular protein function (Kim, Y.H. et al. (1999) *Proc. Natl.*
- 25   *Acad. Sci. USA* 96:12350-12355).

The human h-warts protein, a homolog of *Drosophila* warts tumor suppressor gene, maps to chromosome 6q24-25.1. It has a serine/threonine kinase domain and is localized to centrosomes in interphase cells. It is involved in mitosis and functions as a component of the mitotic apparatus (Nishiyama, Y. et al. (1999) *FEBS Lett.* 459:159-165).

30   **Calcium-Calmodulin Dependent Protein Kinases**

Calcium-calmodulin dependent (CaM) kinases are involved in regulation of smooth muscle contraction, glycogen breakdown (phosphorylase kinase), and neurotransmission (CaM kinase I and CaM kinase II). CaM dependent protein kinases are activated by calmodulin, an intracellular calcium

receptor, in response to the concentration of free calcium in the cell. Many CaM kinases are also activated by phosphorylation. Some CaM kinases are also activated by autophosphorylation or by other regulatory kinases. CaM kinase I phosphorylates a variety of substrates including the neurotransmitter-related proteins synapsin I and II, the gene transcription regulator, CREB, and the cystic fibrosis conductance regulator protein, CFTR (Haribabu, B. et al. (1995) EMBO J. 14:3679-3686). CaM kinase II also phosphorylates synapsin at different sites and controls the synthesis of catecholamines in the brain through phosphorylation and activation of tyrosine hydroxylase. CaM kinase II controls the synthesis of catecholamines and serotonin, through phosphorylation/activation of tyrosine hydroxylase and tryptophan hydroxylase, respectively (Fujisawa, H. (1990) BioEssays 12:27-29). The mRNA encoding a calmodulin-binding protein kinase-like protein was found to be enriched in mammalian forebrain. This protein is associated with vesicles in both axons and dendrites and accumulates largely postnatally. The amino acid sequence of this protein is similar to CaM-dependent STKs, and the protein binds calmodulin in the presence of calcium (Godbout, M. et al. (1994) J. Neurosci. 14:1-13). CaM kinase II phosphorylates the C terminal domain of dystrophin to inhibit its binding to alpha-syntrophin (Madhavan, R. and Jarrett, H.W. (1999) Biochim. Biophys. Acta 1434:260-274). Dystrophin and dystrophin-associated proteins including syntrophins are expressed in skeletal, cardiac, and smooth muscles and in the peripheral and central nervous systems (Ueda, H. et al. (2000) Histol. Histopathol. 15:753-760).

#### Mitogen-Activated Protein Kinases

The mitogen-activated protein kinases (MAP), which mediate signal transduction from the cell surface to the nucleus via phosphorylation cascades, are another STK family that regulates intracellular signaling pathways. Several subgroups have been identified, and each manifests different substrate specificities and responds to distinct extracellular stimuli (Egan, S.E. and R.A. Weinberg (1993) Nature 365:781-783). There are three kinase modules comprising the MAP kinase cascade: MAPK (MAP), MAPK kinase (MAP2K, MAPKK, or MKK), and MKK kinase (MAP3K, MAPKKK, OR MEKK) (Wang, X.S. et al (1998) Biochem. Biophys. Res. Commun. 253:33-37). The extracellular-regulated kinase (ERK) pathway is activated by growth factors and mitogens, for example, epidermal growth factor (EGF), ultraviolet light, hyperosmolar medium, heat shock, or endotoxic lipopolysaccharide (LPS). The closely related though distinct parallel pathways, the c-Jun N-terminal kinase (JNK), or stress-activated kinase (SAPK) pathway, and the p38 kinase pathway are activated by stress stimuli and proinflammatory cytokines such as tumor necrosis factor (TNF) and interleukin-1 (IL-1). Altered MAP kinase expression is implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and

development.. MAP kinase signaling pathways are present in mammalian cells as well as in yeast.

### Cyclin-Dependent Protein Kinases

The cyclin-dependent protein kinases (CDKs) are STKs that control the progression of cells through the cell cycle. The entry and exit of a cell from mitosis are regulated by the synthesis and  
5 destruction of a family of activating proteins called cyclins. Cyclins are small regulatory proteins that bind to and activate CDKs, which then phosphorylate and activate selected proteins involved in the mitotic process. CDKs are unique in that they require multiple inputs to become activated. In addition to cyclin binding, CDK activation requires the phosphorylation of a specific threonine residue and the dephosphorylation of a specific tyrosine residue on the CDK.

10 Another family of STKs associated with the cell cycle are the NIMA (never in mitosis)-related kinases (Neks). Both CDKs and Neks are involved in duplication, maturation, and separation of the microtubule organizing center, the centrosome, in animal cells (Fry, A.M. et al. (1998) EMBO J. 17:470-481)..

### Checkpoint and Cell Cycle Kinases

15 In the process of cell division, the order and timing of cell cycle transitions are under control of cell cycle checkpoints, which ensure that critical events such as DNA replication and chromosome segregation are carried out with precision. If DNA is damaged, e.g. by radiation, a checkpoint pathway is activated that arrests the cell cycle to provide time for repair. If the damage is extensive, apoptosis is induced. In the absence of such checkpoints, the damaged DNA is inherited by aberrant  
20 cells which may cause proliferative disorders such as cancer. Protein kinases play an important role in this process. For example, a specific kinase, checkpoint kinase 1 (Chk1), has been identified in yeast and mammals, and is activated by DNA damage in yeast. Activation of Chk1 leads to the arrest of the cell at the G2/M transition (Sanchez, Y. et al. (1997) Science 277:1497-1501). Specifically, Chk1 phosphorylates the cell division cycle phosphatase CDC25, inhibiting its normal function which is  
25 to dephosphorylate and activate the cyclin-dependent kinase Cdc2. Cdc2 activation controls the entry of cells into mitosis (Peng, C.-Y. et al. (1997) Science 277:1501-1505). Thus, activation of Chk1 prevents the damaged cell from entering mitosis. A deficiency in a checkpoint kinase, such as Chk1, may also contribute to cancer by failure to arrest cells with damaged DNA at other checkpoints such as G2/M.

### 30 Proliferation-Related Kinases

Proliferation-related kinase is a serum/cytokine inducible STK that is involved in regulation of the cell cycle and cell proliferation in human megakaryocytic cells (Li, B. et al. (1996) J. Biol. Chem. 271:19402-19408). Proliferation-related kinase is related to the polo (derived from Drosophila polo

gene) family of STKs implicated in cell division. Proliferation-related kinase is downregulated in lung tumor tissue and may be a proto-oncogene whose deregulated expression in normal tissue leads to oncogenic transformation.

#### 5'-AMP-activated protein kinase

5 A ligand-activated STK protein kinase is 5'-AMP-activated protein kinase (AMPK) (Gao, G. et al. (1996) J. Biol Chem. 271:8675-8681). Mammalian AMPK is a regulator of fatty acid and sterol synthesis through phosphorylation of the enzymes acetyl-CoA carboxylase and hydroxymethylglutaryl-CoA reductase and mediates responses of these pathways to cellular stresses such as heat shock and depletion of glucose and ATP. AMPK is a heterotrimeric complex comprised  
10 of a catalytic alpha subunit and two non-catalytic beta and gamma subunits that are believed to regulate the activity of the alpha subunit. Three isoforms of the gamma subunit,  $\gamma 1$ ,  $\gamma 2$ , and  $\gamma 3$  have been identified (Cheung, P.C. et al. (2000) Biochem. J. 346:659-669). The sensitivity of AMPK to AMP depends on the particular gamma isoform present. Subunits of AMPK have a much wider distribution in non-lipogenic tissues such as brain, heart, spleen, and lung than expected. This  
15 distribution suggests that its role may extend beyond regulation of lipid metabolism alone.

#### Kinases in Apoptosis

Apoptosis is a highly regulated signaling pathway leading to cell death that plays a crucial role in tissue development and homeostasis. Deregulation of this process is associated with the pathogenesis of a number of diseases including autoimmune diseases, neurodegenerative disorders,  
20 and cancer. Various STKs play key roles in this process. ZIP kinase is an STK containing a C-terminal leucine zipper domain in addition to its N-terminal protein kinase domain. This C-terminal domain appears to mediate homodimerization and activation of the kinase as well as interactions with transcription factors such as activating transcription factor, ATF4, a member of the cyclic-AMP responsive element binding protein (ATF/CREB) family of transcriptional factors (Sanjo, H. et al.  
25 (1998) J. Biol. Chem. 273:29066-29071). DRAK1 and DRAK2 are STKs that share homology with the death-associated protein kinases (DAP kinases), known to function in interferon- $\gamma$  induced apoptosis (Sanjo et al., *supra*). Like ZIP kinase, DAP kinases contain a C-terminal protein-protein interaction domain, in the form of ankyrin repeats, in addition to the N-terminal kinase domain. ZIP, DAP, and DRAK kinases induce morphological changes associated with apoptosis when transfected  
30 into NIH3T3 cells (Sanjo et al., *supra*). However, deletion of either the N-terminal kinase catalytic domain or the C-terminal domain of these proteins abolishes apoptosis activity, indicating that in addition to the kinase activity, activity in the C-terminal domain is also necessary for apoptosis, possibly as an interacting domain with a regulator or a specific substrate.

RICK is another STK recently identified as mediating a specific apoptotic pathway involving the death receptor, CD95 (Inohara, N. et al. (1998) J. Biol. Chem. 273:12296-12300). CD95 is a member of the tumor necrosis factor receptor superfamily and plays a critical role in the regulation and homeostasis of the immune system (Nagata, S. (1997) Cell 88:355-365). The CD95 receptor signaling pathway involves recruitment of various intracellular molecules to a receptor complex following ligand binding. This process includes recruitment of the cysteine protease caspase-8 which, in turn, activates a caspase cascade leading to cell death. RICK is composed of an N-terminal kinase catalytic domain and a C-terminal "caspase-recruitment" domain that interacts with caspase-like domains, indicating that RICK plays a role in the recruitment of caspase-8. This interpretation is supported by the fact that the expression of RICK in human 293T cells promotes activation of caspase-8 and potentiates the induction of apoptosis by various proteins involved in the CD95 apoptosis pathway (Inohara et al., *supra*).

#### Mitochondrial Protein Kinases

A novel class of eukaryotic kinases, related by sequence to prokaryotic histidine protein kinases, are the mitochondrial protein kinases (MPKs) which seem to have no sequence similarity with other eukaryotic protein kinases. These protein kinases are located exclusively in the mitochondrial matrix space and may have evolved from genes originally present in respiration-dependent bacteria which were endocytosed by primitive eukaryotic cells. MPKs are responsible for phosphorylation and inactivation of the branched-chain alpha-ketoacid dehydrogenase and pyruvate dehydrogenase complexes (Harris, R.A. et al. (1995) Adv. Enzyme Regul. 34:147-162). Five MPKs have been identified. Four members correspond to pyruvate dehydrogenase kinase isozymes, regulating the activity of the pyruvate dehydrogenase complex, which is an important regulatory enzyme at the interface between glycolysis and the citric acid cycle. The fifth member corresponds to a branched-chain alpha-ketoacid dehydrogenase kinase, important in the regulation of the pathway for the disposal of branched-chain amino acids. (Harris, R.A. et al. (1997) Adv. Enzyme Regul. 37:271-293). Both starvation and the diabetic state are known to result in a great increase in the activity of the pyruvate dehydrogenase kinase in the liver, heart and muscle of the rat. This increase contributes in both disease states to the phosphorylation and inactivation of the pyruvate dehydrogenase complex and conservation of pyruvate and lactate for gluconeogenesis (Harris (1995) *supra*).

#### KINASES WITH NON-PROTEIN SUBSTRATES

##### Lipid and Inositol kinases

Lipid kinases phosphorylate hydroxyl residues on lipid head groups. A family of kinases involved in phosphorylation of phosphatidylinositol (PI) has been described, each member

phosphorylating a specific carbon on the inositol ring (Leevers, S.J. et al. (1999) *Curr. Opin. Cell. Biol.* 11:219-225). The phosphorylation of phosphatidylinositol is involved in activation of the protein kinase C signaling pathway. The inositol phospholipids (phosphoinositides) intracellular signaling pathway begins with binding of a signaling molecule to a G-protein linked receptor in the plasma membrane.

5 This leads to the phosphorylation of phosphatidylinositol (PI) residues on the inner side of the plasma membrane by inositol kinases, thus converting PI residues to the biphosphate state ( $PIP_2$ ).  $PIP_2$  is then cleaved into inositol triphosphate ( $IP_3$ ) and diacylglycerol. These two products act as mediators for separate signaling pathways. Cellular responses that are mediated by these pathways are glycogen breakdown in the liver in response to vasopressin, smooth muscle contraction in response to

10 acetylcholine, and thrombin-induced platelet aggregation.

PI 3-kinase (PI3K), which phosphorylates the D3 position of PI and its derivatives, has a central role in growth factor signal cascades involved in cell growth, differentiation, and metabolism. PI3K is a heterodimer consisting of an adapter subunit and a catalytic subunit. The adapter subunit acts as a scaffolding protein, interacting with specific tyrosine-phosphorylated proteins, lipid moieties,

15 and other cytosolic factors. When the adapter subunit binds tyrosine phosphorylated targets, such as the insulin responsive substrate (IRS)-1, the catalytic subunit is activated and converts PI (4,5) biphosphate ( $PIP_2$ ) to PI (3,4,5)  $P_3$  ( $PIP_3$ ).  $PIP_3$  then activates a number of other proteins, including PKA, protein kinase B (PKB), protein kinase C (PKC), glycogen synthase kinase (GSK)-3, and p70 ribosomal s6 kinase. PI3K also interacts directly with the cytoskeletal organizing proteins; Rac, rho,

20 and cdc42 (Shepherd, P.R. et al. (1998) *Biochem. J.* 333:471-490). Animal models for diabetes, such as *obese* and *fat* mice, have altered PI3K adapter subunit levels. Specific mutations in the adapter subunit have also been found in an insulin-resistant Danish population, suggesting a role for PI3K in type-2 diabetes (Shepard, *supra*).

An example of lipid kinase phosphorylation activity is the phosphorylation of

25 D-erythro-sphingosine to the sphingolipid metabolite, sphingosine-1-phosphate (SPP). SPP has emerged as a novel lipid second-messenger with both extracellular and intracellular actions (Kohama, T. et al. (1998) *J. Biol. Chem.* 273:23722-23728). Extracellularly, SPP is a ligand for the G-protein coupled receptor EDG-1 (endothelial-derived, G-protein coupled receptor). Intracellularly, SPP regulates cell growth, survival, motility, and cytoskeletal changes. SPP levels are regulated by

30 sphingosine kinases that specifically phosphorylate D-erythro-sphingosine to SPP. The importance of sphingosine kinase in cell signaling is indicated by the fact that various stimuli, including platelet-derived growth factor (PDGF), nerve growth factor, and activation of protein kinase C, increase cellular levels of SPP by activation of sphingosine kinase, and the fact that competitive

inhibitors of the enzyme selectively inhibit cell proliferation induced by PDGF (Kohama et al., supra).

### Purine Nucleotide Kinases

The purine nucleotide kinases, adenylate kinase (ATP:AMP phosphotransferase, or AdK) and guanylate kinase (ATP:GMP phosphotransferase, or GuK) play a key role in nucleotide metabolism and are crucial to the synthesis and regulation of cellular levels of ATP and GTP, respectively. These two molecules are precursors in DNA and RNA synthesis in growing cells and provide the primary source of biochemical energy in cells (ATP), and signal transduction pathways (GTP). Inhibition of various steps in the synthesis of these two molecules has been the basis of many antiproliferative drugs for cancer and antiviral therapy (Pillwein, K. et al. (1990) Cancer Res. 50:1576-1579).

AdK is found in almost all cell types and is especially abundant in cells having high rates of ATP synthesis and utilization such as skeletal muscle. In these cells AdK is physically associated with mitochondria and myofibrils, the subcellular structures that are involved in energy production and utilization, respectively. Recent studies have demonstrated a major function for AdK in transferring high energy phosphoryls from metabolic processes generating ATP to cellular components consuming ATP (Zelevnikar, R.J. et al. (1995) J. Biol. Chem. 270:7311-7319). Thus AdK may have a pivotal role in maintaining energy production in cells, particularly those having a high rate of growth or metabolism such as cancer cells, and may provide a target for suppression of its activity in order to treat certain cancers. Alternatively, reduced AdK activity may be a source of various metabolic, muscle-energy disorders that can result in cardiac or respiratory failure and may be treatable by increasing AdK activity.

GuK, in addition to providing a key step in the synthesis of GTP for RNA and DNA synthesis, also fulfills an essential function in signal transduction pathways of cells through the regulation of GDP and GTP. Specifically, GTP binding to membrane associated G proteins mediates the activation of cell receptors, subsequent intracellular activation of adenylyl cyclase, and production of the second messenger, cyclic AMP. GDP binding to G proteins inhibits these processes. GDP and GTP levels also control the activity of certain oncogenic proteins such as p21<sup>ras</sup> known to be involved in control of cell proliferation and oncogenesis (Bos, J.L. (1989) Cancer Res. 49:4682-4689). High ratios of GTP:GDP caused by suppression of GuK cause activation of p21<sup>ras</sup> and promote oncogenesis. Increasing GuK activity to increase levels of GDP and reduce the GTP:GDP ratio may provide a therapeutic strategy to reverse oncogenesis.

GuK is an important enzyme in the phosphorylation and activation of certain antiviral drugs useful in the treatment of herpes virus infections. These drugs include the guanine homologs acyclovir and buciclovir (Miller, W.H. and R.L. Miller (1980) J. Biol. Chem. 255:7204-7207; Stenberg, K. et al.



(1986) J. Biol. Chem. 261:2134-2139). Increasing GuK activity in infected cells may provide a therapeutic strategy for augmenting the effectiveness of these drugs and possibly for reducing the necessary dosages of the drugs.

### Pyrimidine Kinases

5        The pyrimidine kinases are deoxycytidine kinase and thymidine kinase 1 and 2. Deoxycytidine kinase is located in the nucleus, and thymidine kinase 1 and 2 are found in the cytosol (Johansson, M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11941-11945). Phosphorylation of deoxyribonucleosides by pyrimidine kinases provides an alternative pathway for de novo synthesis of DNA precursors. The role of pyrimidine kinases, like purine kinases, in phosphorylation is critical to the activation of several  
10        chemotherapeutically important nucleoside analogues (Arner E.S. and S. Eriksson (1995) Pharmacol. Ther. 67:155-186).

### PHOSPHATASES

Protein phosphatases are generally characterized as either serine/threonine- or tyrosine-specific based on their preferred phospho-amino acid substrate. However, some phosphatases (DSPs,  
15        for dual specificity phosphatases) can act on phosphorylated tyrosine, serine, or threonine residues. The protein serine/threonine phosphatases (PSPs) are important regulators of many cAMP-mediated hormone responses in cells. Protein tyrosine phosphatases (PTPs) play a significant role in cell cycle and cell signaling processes. Another family of phosphatases is the acid phosphatase or histidine acid phosphatase (HAP) family whose members hydrolyze phosphate esters at acidic pH conditions.

20        PSPs are found in the cytosol, nucleus, and mitochondria and in association with cytoskeletal and membranous structures in most tissues, especially the brain. Some PSPs require divalent cations, such as  $\text{Ca}^{2+}$  or  $\text{Mn}^{2+}$ , for activity. PSPs play important roles in glycogen metabolism, muscle contraction, protein synthesis, T cell function, neuronal activity, oocyte maturation, and hepatic metabolism (reviewed in Cohen, P. (1989) Annu. Rev. Biochem. 58:453-508). PSPs can be separated  
25        into two classes. The PPP class includes PP1, PP2A, PP2B/calcineurin, PP4, PP5, PP6, and PP7. Members of this class are composed of a homologous catalytic subunit bearing a very highly conserved signature sequence, coupled with one or more regulatory subunits (PROSITE PDOC00115). Further interactions with scaffold and anchoring molecules determine the intracellular localization of PSPs and substrate specificity. The PPM class consists of several closely related  
30        isoforms of PP2C and is evolutionarily unrelated to the PPP class.

PP1 dephosphorylates many of the proteins phosphorylated by cyclic AMP-dependent protein kinase (PKA) and is an important regulator of many cAMP-mediated hormone responses in cells. A number of isoforms have been identified, with the alpha and beta forms being produced by alternative

splicing of the same gene. Both ubiquitous and tissue-specific targeting proteins for PP1 have been identified. In the brain, inhibition of PP1 activity by the dopamine and adenosine 3',5'-monophosphate-regulated phosphoprotein of 32kDa (DARPP-32) is necessary for normal dopamine response in neostriatal neurons (reviewed in Price, N.E. and M.C. Mumby (1999) *Curr. Opin. Neurobiol.* 9:336-342). PP1, along with PP2A, has been shown to limit motility in microvascular endothelial cells, suggesting a role for PSPs in the inhibition of angiogenesis (Gabel, S. et al. (1999) *Otolaryngol. Head Neck Surg.* 121:463-468).

PP2A is the main serine/threonine phosphatase. The core PP2A enzyme consists of a single 36 kDa catalytic subunit (C) associated with a 65 kDa scaffold subunit (A), whose role is to recruit additional regulatory subunits (B). Three gene families encoding B subunits are known (PR55, PR61, and PR72), each of which contain multiple isoforms, and additional families may exist (Millward, T.A et al. (1999) *Trends Biosci.* 24:186-191). These "B-type" subunits are cell type- and tissue-specific and determine the substrate specificity, enzymatic activity, and subcellular localization of the holoenzyme. The PR55 family is highly conserved and bears a conserved motif (PROSITE PDOC00785). PR55 increases PP2A activity toward mitogen-activated protein kinase (MAPK) and MAPK kinase (MEK). PP2A dephosphorylates the MAPK active site, inhibiting the cell's entry into mitosis. Several proteins can compete with PR55 for PP2A core enzyme binding, including the CKII kinase catalytic subunit, polyomavirus middle and small T antigens, and SV40 small t antigen. Viruses may use this mechanism to commandeer PP2A and stimulate progression of the cell through the cell cycle (Pallas, D.C. et al. (1992) *J. Virol.* 66:886-893). Altered MAP kinase expression is also implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development. PP2A, in fact, can dephosphorylate and modulate the activities of more than 30 protein kinases in vitro, and other evidence suggests that the same is true in vivo for such kinases as PKB, PKC, the calmodulin-dependent kinases, ERK family MAP kinases, cyclin-dependent kinases, and the I $\kappa$ B kinases (reviewed in Millward et al., supra). PP2A is itself a substrate for CKI and CKII kinases, and can be stimulated by polycationic macromolecules. A PP2A-like phosphatase is necessary to maintain the G1 phase destruction of mammalian cyclins A and B (Bastians, H. et al. (1999) *Mol. Biol. Cell* 10:3927-3941). PP2A is a major activity in the brain and is implicated in regulating neurofilament stability and normal neural function, particularly the phosphorylation of the microtubule-associated protein tau. Hyperphosphorylation of tau has been proposed to lead to the neuronal degeneration seen in Alzheimer's disease (reviewed in Price and Mumby, supra).

PP2B, or calcineurin, is a Ca<sup>2+</sup>-activated dimeric phosphatase and is particularly abundant in

the brain. It consists of catalytic and regulatory subunits, and is activated by the binding of the calcium/calmodulin complex. Calcineurin is the target of the immunosuppressant drugs cyclosporine and FK506. Along with other cellular factors, these drugs interact with calcineurin and inhibit phosphatase activity. In T cells, this blocks the calcium dependent activation of the NF-AT family of transcription factors, leading to immunosuppression. This family is widely distributed, and it is likely that calcineurin regulates gene expression in other tissues as well. In neurons, calcineurin modulates functions which range from the inhibition of neurotransmitter release to desensitization of postsynaptic NMDA-receptor coupled calcium channels to long term memory (reviewed in Price and Mumby, supra).

Other members of the PPP class have recently been identified (Cohen, P.T. (1997) Trends Biochem. Sci. 22:245-251). One of them, PP5, contains regulatory domains with tetratricopeptide repeats. It can be activated by polyunsaturated fatty acids and anionic phospholipids in vitro and appears to be involved in a number of signaling pathways, including those controlled by atrial natriuretic peptide or steroid hormones (reviewed in Andreeva, A.V. and M.A. Kutuzov (1999) Cell Signal. 11:555-562).

PP2C is a ~42kDa monomer with broad substrate specificity and is dependent on divalent cations (mainly  $Mn^{2+}$  or  $Mg^{2+}$ ) for its activity. PP2C proteins share a conserved N-terminal region with an invariant DGH motif, which contains an aspartate residue involved in cation binding (PROSITE PDOC00792). Targeting proteins and mechanisms regulating PP2C activity have not been identified. PP2C has been shown to inhibit the stress-responsive p38 and Jun kinase (JNK) pathways (Takekawa, M. et al. (1998) EMBO J. 17:4744-4752).

In contrast to PSPs, tyrosine-specific phosphatases (PTPs) are generally monomeric proteins of very diverse size (from 20kDa to greater than 100kDa) and structure that function primarily in the transduction of signals across the plasma membrane. PTPs are categorized as either soluble phosphatases or transmembrane receptor proteins that contain a phosphatase domain. All PTPs share a conserved catalytic domain of about 300 amino acids which contains the active site. The active site consensus sequence includes a cysteine residue which executes a nucleophilic attack on the phosphate moiety during catalysis (Neel, B.G. and N.K. Tonks (1997) Curr. Opin. Cell Biol. 9:193-204). Receptor PTPs are made up of an N-terminal extracellular domain of variable length, a transmembrane region, and a cytoplasmic region that generally contains two copies of the catalytic domain. Although only the first copy seems to have enzymatic activity, the second copy apparently affects the substrate specificity of the first. The extracellular domains of some receptor PTPs contain fibronectin-like repeats, immunoglobulin-like domains, MAM domains (an extracellular motif likely to

have an adhesive function), or carbonic anhydrase-like domains (PROSITE PDOC 00323). This wide variety of structural motifs accounts for the diversity in size and specificity of PTPs.

PTPs play important roles in biological processes such as cell adhesion, lymphocyte activation, and cell proliferation. PTPs  $\mu$  and  $\kappa$  are involved in cell-cell contacts, perhaps regulating  
5 cadherin/catenin function. A number of PTPs affect cell spreading, focal adhesions, and cell motility, most of them via the integrin/tyrosine kinase signaling pathway (reviewed in Neel and Tonks, supra). CD45 phosphatases regulate signal transduction and lymphocyte activation (Ledbetter, J.A. et al. (1988) Proc. Natl. Acad. Sci. USA 85:8628-8632). Soluble PTPs containing Src-homology-2 domains have been identified (SHPs), suggesting that these molecules might interact with receptor tyrosine  
10 kinases. SHP-1 regulates cytokine receptor signaling by controlling the Janus family PTKs in hematopoietic cells, as well as signaling by the T-cell receptor and c-Kit (reviewed in Neel and Tonks, supra). M-phase inducer phosphatase plays a key role in the induction of mitosis by dephosphorylating and activating the PTK CDC2, leading to cell division (Sadhu, K. et al. (1990) Proc. Natl. Acad. Sci. USA 87:5139-5143). In addition, the genes encoding at least eight PTPs have been mapped to  
15 chromosomal regions that are translocated or rearranged in various neoplastic conditions, including lymphoma, small cell lung carcinoma, leukemia, adenocarcinoma, and neuroblastoma (reviewed in Charbonneau, H. and N.K. Tonks (1992) Annu. Rev. Cell Biol. 8:463-493). The PTP enzyme active site comprises the consensus sequence of the MTM1 gene family. The MTM1 gene is responsible for X-linked recessive myotubular myopathy, a congenital muscle disorder that has been linked to Xq28  
20 (Kioschis, P. et al., (1998) Genomics 54:256-266). Many PTKs are encoded by oncogenes, and it is well known that oncogenesis is often accompanied by increased tyrosine phosphorylation activity. It is therefore possible that PTPs may serve to prevent or reverse cell transformation and the growth of various cancers by controlling the levels of tyrosine phosphorylation in cells. This is supported by studies showing that overexpression of PTP can suppress transformation in cells and that specific  
25 inhibition of PTP can enhance cell transformation (Charbonneau and Tonks, supra).

Dual specificity phosphatases (DSPs) are structurally more similar to the PTPs than the PSPs. DSPs bear an extended PTP active site motif with an additional 7 amino acid residues. DSPs are primarily associated with cell proliferation and include the cell cycle regulators cdc25A, B, and C. The phosphatases DUSP1 and DUSP2 inactivate the MAPK family members ERK (extracellular  
30 signal-regulated kinase), JNK (c-Jun N-terminal kinase), and p38 on both tyrosine and threonine residues (PROSITE PDOC 00323, supra). In the activated state, these kinases have been implicated in neuronal differentiation, proliferation, oncogenic transformation, platelet aggregation, and apoptosis. Thus, DSPs are necessary for proper regulation of these processes (Muda, M. et al. (1996) J. Biol.

Chem. 271:27205-27208). The tumor suppressor PTEN is a DSP that also shows lipid phosphatase activity. It seems to negatively regulate interactions with the extracellular matrix and maintains sensitivity to apoptosis. PTEN has been implicated in the prevention of angiogenesis (Giri, D. and M. Ittmann (1999) Hum. Pathol. 30:419-424) and abnormalities in its expression are associated with  
5 numerous cancers (reviewed in Tamura, M. et al. (1999) J. Natl. Cancer Inst. 91:1820-1828).

Histidine acid phosphatase (HAP; EXPASY EC 3.1.3.2), also known as acid phosphatase, hydrolyzes a wide spectrum of substrates including alkyl, aryl, and acyl orthophosphate monoesters and phosphorylated proteins at low pH. HAPs share two regions of conserved sequences, each centered around a histidine residue which is involved in catalytic activity. Members of the HAP  
10 family include lysosomal acid phosphatase (LAP) and prostatic acid phosphatase (PAP), both sensitive to inhibition by L-tartrate (PROSITE PDOC00538).

LAP, an orthophosphoric monoester of the endosomal/lysosomal compartment is a housekeeping gene whose enzymatic activity has been detected in all tissues examined (Geier, C. et al. (1989) Eur. J. Biochem. 183:611-616). LAP-deficient mice have progressive skeletal disorder and  
15 an increased disposition toward generalized seizures (Saftig, P. et al. (1997) J. Biol. Chem. 272:18628-18635). LAP-deficient patients were found to have the following clinical features: intermittent vomiting, hypotonia, lethargy, opisthotonos, terminal bleeding, seizures, and death in early infancy (Online Mendelian Inheritance in Man (OMIM) \*200950).

PAP, a prostate epithelium-specific differentiation antigen produced by the prostate gland, has  
20 been used to diagnose and stage prostate cancer. In prostate carcinomas, the enzymatic activity of PAP was shown to be decreased compared with normal or benign prostate hypertrophy cells (Foti, A.G. et al. (1977) Cancer Res. 37:4120-4124). Two forms of PAP have been identified, secreted and intracellular. Mature secreted PAP is detected in the seminal fluid and is active as a glycosylated homodimer with a molecular weight of approximately 100-kilodalton. Intracellular PAP is found to  
25 exhibit endogenous phosphotyrosyl protein phosphatase activity and is involved in regulating prostate cell growth (Meng, T.C. and M.F. Lin (1998) J. Biol. Chem. 34:22096-22104).

LAP, an orthophosphoric monoester of the endosomal/lysosomal compartment is a housekeeping gene whose enzymatic activity has been detected in all tissues examined (Geier, C. et al. (1989) Eur. J. Biochem. 183:611-616). LAP-deficient mice have progressive skeletal disorder and  
30 an increased disposition toward generalized seizures (Saftig, P. et al. (1997) J. Biol. Chem. 272:18628-18635). LAP-deficient patients were found to have the following clinical features: intermittent vomiting, hypotonia, lethargy, opisthotonos, terminal bleeding, seizures, and death in early infancy (Online Mendelian Inheritance in Man (OMIM) \*200950).

PAP, a prostate epithelium-specific differentiation antigen produced by the prostate gland, has been used to diagnose and stage prostate cancer. In prostate carcinomas, the enzymatic activity of PAP was shown to be decreased compared with normal or benign prostate hypertrophy cells (Foti, A.G. et al. (1977) Cancer Res. 37:4120-4124). Two forms of PAP have been identified, secreted and  
5 intracellular. Mature secreted PAP is detected in the seminal fluid and is active as a glycosylated homodimer with a molecular weight of approximately 100-kilodalton. Intracellular PAP is found to exhibit endogenous phosphotyrosyl protein phosphatase activity and is involved in regulating prostate cell growth (Meng, T.C. and M.F. Lin (1998) J. Biol. Chem. 34:22096-22104).

Synaptojanin, a polyphosphoinositide phosphatase, dephosphorylates phosphoinositides at  
10 positions 3, 4 and 5 of the inositol ring. Synaptojanin is a major presynaptic protein found at clathrin-coated endocytic intermediates in nerve terminals, and binds the clathrin coat-associated protein, EPS15. This binding is mediated by the C-terminal region of synaptojanin-170, which has 3 Asp-Pro-Phe amino acid repeats. Further, this 3 residue repeat had been found to be the binding site for the BH domains of EPS15 (Haffner, C. et al. (1997) FEBS Lett. 419:175-180). Additionally, synaptojanin  
15 may potentially regulate interactions of endocytic proteins with the plasma membrane, and be involved in synaptic vesicle recycling (Brodin, L. et al. (2000) Curr. Opin. Neurobiol. 10:312-320). Studies in mice with a targeted disruption in the synaptojanin 1 gene (Synj1) were shown to support coat formation of endocytic vesicles more effectively than was seen in wild-type mice, suggesting that Synj1 can act as a negative regulator of membrane-coat protein interactions. These findings provide  
20 genetic evidence for a crucial role of phosphoinositide metabolism in synaptic vesicle recycling (Cremona, O. et al. (1999) Cell 99:179-188).

#### Expression profiling

Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression  
25 of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

30 The discovery of new kinases and phosphatases, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers, and in the

assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of kinases and phosphatases.

### SUMMARY OF THE INVENTION

5       The invention features purified polypeptides, kinases and phosphatases, referred to collectively as "KPP" and individually as "KPP-1," "KPP-2," "KPP-3," "KPP-4," "KPP-5," "KPP-6," "KPP-7," "KPP-8," "KPP-9," "KPP-10," "KPP-11," "KPP-12," "KPP-13," "KPP-14," and "KPP-15." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 99% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-15.

15       The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 99% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-15. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:16-30.

25       Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 99% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the

invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least  
5 99% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is  
10 transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally  
15 occurring amino acid sequence at least 99% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.

20 The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to  
25 the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ  
30 ID NO:16-30, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the



sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 99% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 99% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an

immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a  
5 pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an  
10 amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 99% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the  
15 group consisting of SEQ ID NO:1-15. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of  
20 treating a disease or condition associated with overexpression of functional KPP, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 99% identical to an amino acid sequence selected  
25 from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test  
30 compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a

naturally occurring amino acid sequence at least 99% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment

of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

### BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

### DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a

reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs.

5 Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the  
10 invention is not entitled to antedate such disclosure by virtue of prior invention.

### DEFINITIONS

"KPP" refers to the amino acid sequences of substantially purified KPP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

15 The term "agonist" refers to a molecule which intensifies or mimics the biological activity of KPP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of KPP either by directly interacting with KPP or by acting on components of the biological pathway in which KPP participates.

An "allelic variant" is an alternative form of the gene encoding KPP. Allelic variants may  
20 result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times  
25 in a given sequence.

"Altered" nucleic acid sequences encoding KPP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as KPP or a polypeptide with at least one functional characteristic of KPP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of  
30 the polynucleotide encoding KPP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding KPP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent KPP. Deliberate

amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of KPP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of KPP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of KPP either by directly interacting with KPP or by acting on components of the biological pathway in which KPP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind KPP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on

the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an in vitro evolutionary process (e.g., SELEX  
5 (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a  
10 ribonucleotide may be replaced by 2'-F or 2'-NH<sub>2</sub>), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

15 The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed  
20 nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as  
25 phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring  
30 nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical

functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic KPP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding KPP or fragments of KPP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

|    | Original Residue | Conservative Substitution |
|----|------------------|---------------------------|
|    | Ala              | Gly, Ser                  |
|    | Arg              | His, Lys                  |
| 30 | Asn              | Asp, Gln, His             |
|    | Asp              | Asn, Glu                  |
|    | Cys              | Ala, Ser                  |
|    | Gln              | Asn, Glu, His             |
|    | Glu              | Asp, Gln, His             |
| 35 | Gly              | Ala                       |
|    | His              | Asn, Arg, Gln, Glu        |



|    |     |                         |
|----|-----|-------------------------|
|    | Ile | Leu, Val                |
|    | Leu | Ile, Val                |
|    | Lys | Arg, Gln, Glu           |
|    | Met | Leu, Ile                |
| 5  | Phe | His, Met, Leu, Trp, Tyr |
|    | Ser | Cys, Thr                |
|    | Thr | Ser, Val                |
|    | Trp | Phe, Tyr                |
|    | Tyr | His, Phe, Trp           |
| 10 | Val | Ile, Leu, Thr           |

---

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of KPP or the polynucleotide encoding KPP which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment

used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:16-30 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:16-30, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:16-30 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:16-30 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:16-30 and the region of SEQ ID NO:16-30 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-15 is encoded by a fragment of SEQ ID NO:16-30. A fragment of SEQ ID NO:1-15 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-15. For example, a fragment of SEQ ID NO:1-15 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-15. The precise length of a fragment of SEQ ID NO:1-15 and the region of SEQ ID NO:1-15 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default

parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Reward for match: 1*

*Penalty for mismatch: -2*

*Open Gap: 5 and Extension Gap: 2 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 11*

*Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous

nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode  
5 similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a  
10 standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default  
15 parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by  
20 CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

25        *Matrix: BLOSUM62*  
         *Open Gap: 11 and Extension Gap: 1 penalties*  
         *Gap x drop-off: 50*  
         *Expect: 10*  
         *Word Size: 3*  
30        *Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for

instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

5       “Human artificial chromosomes” (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term “humanized antibody” refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely  
10 resembles a human antibody, and still retains its original binding ability.

“Hybridization” refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized  
15 after the “washing” step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be  
20 varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about  
25 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY;  
30 specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may

be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200  $\mu\text{g/ml}$ . Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g.,  $C_0t$  or  $R_0t$  analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of KPP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of KPP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of KPP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of KPP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or

synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

“Operably linked” refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably  
5 linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

“Peptide nucleic acid” (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of  
10 amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

“Post-translational modification” of an KPP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the  
15 art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of KPP.

“Probe” refers to nucleic acid sequences encoding KPP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule.  
20 Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. “Primers” are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

25 Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the  
30 specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular

Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a



recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is  
5 expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

10 "Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear  
15 sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing KPP, nucleic acids encoding KPP, or fragments thereof may comprise a bodily fluid; an extract from a cell,  
20 chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure  
25 of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are  
30 removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides

by different amino acid residues or nucleotides, respectively.

“Substrate” refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A “transcript image” or “expression profile” refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

“Transformation” describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term “transformed cells” includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A “transgenic organism,” as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In one alternative, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), *supra*.

A “variant” of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the “BLAST 2 Sequences” tool Version 2.0.9 (May-07-

1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an  
5 "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to  
10 another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a  
15 propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-  
20 1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

## THE INVENTION

25 The invention is based on the discovery of new human kinases and phosphatases (KPP), the polynucleotides encoding KPP, and the use of these compositions for the diagnosis, treatment, or prevention of cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide  
30 sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is

denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are kinases and phosphatases. For example, SEQ ID NO:2 is 587 amino acids in length and is 94% identical, from residue M1 to residue Q536, to human 63kDa protein kinase (GenBank ID g23903) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $2.4e-271$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains an eukaryotic protein kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:2 is a MAP kinase isoform P63 protein kinase (SWISS\_PROT:P31152). In another example, SEQ ID NO:5 is 50% identical, from residue G55 to residue V439, to rabbit myosin light chain kinase (GenBank ID g165506) as determined by BLAST. The BLAST probability score is  $5.3e-100$ . SEQ ID NO:5 also contains a eukaryotic protein kinase domain as determined by searching for statistically significant matches in the hidden Markov model

(HMM)-based PFAM database. Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:5 is a human kinase. In yet another example, SEQ ID NO:7 is 53% identical, from residue G33 to residue P1312, to rat SCOP (suprachiasmatic nucleus (SCN) circadian oscillatory protein) (GenBank ID g4884492), a putative 2C-type protein phosphatase, as determined by BLAST. The BLAST probability score is 0.0. SEQ ID NO:7 also contains a protein phosphatase 2C domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database. Additional data obtained from searching the PFAM database, and from BLIMPS and MOTIFS analyses, also provide evidence that SEQ ID NO:7 comprises leucine-rich repeats, associated with protein-protein interactions. In another example, SEQ ID NO:9 is 69% identical, from residue W43 to residue D602, to human mixed lineage kinase MLK1 (GenBank ID g12005724) as determined by BLAST. The BLAST probability score is  $6.9e-250$ . SEQ ID NO:9 also contains an SH3 kinase domain as well as a protein kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database. Data from BLIMPS, MOTIFS, PROFILESCAN, and additional BLAST analyses provide further corroborative evidence that SEQ ID NO:9 is a protein kinase. In another example, SEQ ID NO:10 is 36% identical, from residue R48 to residue N234, to rat adenylate kinase isozyme 1 (GenBank ID g8918488) as determined by BLAST. The BLAST probability score is  $8.1e-33$ . SEQ ID NO:10 also contains an adenylate kinase active site domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database. Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:10 is an adenylate kinase. In a further example, SEQ ID NO:11 is 91% identical, from residue M1 to residue D550, to a human testes-specific putative tyrosine phosphatase (GenBank ID g3549240) as determined by BLAST. The BLAST probability score is  $1.3e-268$ . SEQ ID NO:11 also contains a tyrosine specific protein phosphatases signature domain as determined by searching for statistically significant matches in the PROFILESCAN database of conserved protein family domains. Data from MOTIFS analysis provide further corroborative evidence that SEQ ID NO:11 is a tyrosine phosphatase. In yet another example, SEQ ID NO:12 is 79% identical from residue M1 to residue P1544, and 62% identical from residue P1053 to residue P1547, to *Mus musculus* syntrophin-associated serine threonine protein kinase (GenBank ID g5757703) as determined by BLAST. The BLAST probability score is 0.0 from residue M1 to residue P1544, and  $1.5e-143$  from residue P1053 to residue P1547, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:12 also contains a PDZ membrane associated domain, a protein kinase domain, and a protein kinase C terminal domain as determined by searching for statistically

significant matches in the hidden Markov model (HMM)-based PFAM database. Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:12 is a syntrophin-associated serine threonine kinase. In a further example, SEQ ID NO:13 is 30% identical from residue I354 to residue L498, 40% identical from residue G259 to residue G310, and 31% identical from residue K115 to residue E186, to *Fagus sylvatica* protein phosphatase 2C (GenBank ID g7768151) as determined by BLAST. The BLAST probability score is  $4.4e-12$ . SEQ ID NO:13 also contains a protein phosphatase 2C domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database. Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:13 is a protein phosphatase 2C. In another example, SEQ ID NO:14 is 96% identical, from residue M1 to residue V1036, to *Mus musculus* receptor tyrosine kinase (GenBank ID g1457961) as determined by BLAST. The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:14 also contains an Ephrin receptor ligand binding domain, a SAM domain, a fibronectin type III domain, and a protein kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database. Data from BLIMPS, MOTIFS, and PROFILESCAN analyses and BLAST analyses of the PRODOM and DOMO databases provide further corroborative evidence that SEQ ID NO:14 is a protein tyrosine kinase. In addition, TMAP analysis indicates that SEQ ID NO:14 contains five transmembrane domains. In a further example, SEQ ID NO:15 is 98% identical, from residue M1 to residue G488, to the human AMP-activated protein kinase gamma 3 subunit (GenBank ID g6688201) as determined by BLAST. The BLAST probability score is  $9.3e-261$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:15 also contains four CBS domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database. CBS domains are found in proteins of the AMP-activated protein kinase gamma subunit family. Data from PRODOM and DOMO analyses provide further corroborative evidence that SEQ ID NO:15 is a protein kinase. SEQ ID NO:1, SEQ ID NO:3-4, SEQ ID NO:6, SEQ ID NO:8, and SEQ ID NO:9 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-15 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide

consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:16-30 or that distinguish between SEQ ID NO:16-30 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL\_XXXXXX\_N<sub>1</sub>\_N<sub>2</sub>\_YYYY\_N<sub>3</sub>\_N<sub>4</sub> represents a "stitched" sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYY is the number of the prediction generated by the algorithm, and N<sub>1,2,3,...</sub>, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FLXXXXXX\_gAAAAA\_gBBBBB\_1\_N is a "stretched" sequence, with XXXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, gBBBBB).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The

following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

| Prefix            | Type of analysis and/or examples of programs  |
|-------------------|---|
| GNN, GFG,<br>ENST | Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).                      |
| GBI               | Hand-edited analysis of genomic sequences.  |
| FL                | Stitched or stretched genomic sequences (see Example V).  |
| INCY              | Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript. |

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses KPP variants. A preferred KPP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the KPP amino acid sequence, and which contains at least one functional or structural characteristic of KPP.

The invention also encompasses polynucleotides which encode KPP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:16-30, which encodes KPP. The polynucleotide sequences of SEQ ID NO:16-30, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding KPP. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide



sequence encoding KPP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:16-30 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:16-30. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of KPP.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding KPP. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding KPP, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to the polynucleotide sequence encoding KPP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding KPP. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of KPP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding KPP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring KPP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode KPP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring KPP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding KPP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding KPP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than

transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode KPP and KPP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding KPP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:16-30 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding KPP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids*

Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown  
5 sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using  
10 commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been  
15 size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze  
20 the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire  
25 process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode KPP may be cloned in recombinant DNA molecules that direct expression of KPP, or  
30 fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express KPP.

The nucleotide sequences of the present invention can be engineered using methods generally

known in the art in order to alter KPP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of KPP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding KPP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, KPP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of KPP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid

chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, *supra*, pp. 28-53.)

In order to express a biologically active KPP, the nucleotide sequences encoding KPP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding KPP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding KPP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding KPP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding KPP and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. (See, e.g., Sambrook, J. et al. (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding KPP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, *supra*; Ausubel, *supra*; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA*

91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding KPP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding KPP can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding KPP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of KPP are needed, e.g. for the production of antibodies, vectors which direct high level expression of KPP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of KPP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of KPP. Transcription of sequences encoding KPP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock

promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding KPP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses KPP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of KPP in cell lines is preferred. For example, sequences encoding KPP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *ap<sup>r</sup>* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat*

confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding KPP is inserted within a marker gene sequence, transformed cells containing sequences encoding KPP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding KPP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding KPP and that express KPP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of KPP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on KPP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization



or PCR probes for detecting sequences related to polynucleotides encoding KPP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding KPP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to  
5 synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as  
10 substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding KPP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing  
15 polynucleotides which encode KPP may be designed to contain signal sequences which direct secretion of KPP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation,  
20 lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and  
25 processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding KPP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric KPP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the  
30 screening of peptide libraries for inhibitors of KPP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin

(HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the KPP encoding sequence and the heterologous protein sequence, so that KPP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled KPP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

KPP of the present invention or fragments thereof may be used to screen for compounds that specifically bind to KPP. At least one and up to a plurality of test compounds may be screened for specific binding to KPP. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of KPP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which KPP binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express KPP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing KPP or cell membrane fractions which contain KPP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either KPP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with KPP, either in solution or affixed to a solid support, and detecting the binding of KPP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor.

Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

KPP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of KPP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for KPP activity, wherein KPP is combined with at least one test compound, and the activity of KPP in the presence of a test compound is compared with the activity of KPP in the absence of the test compound. A change in the activity of KPP in the presence of the test compound is indicative of a compound that modulates the activity of KPP. Alternatively, a test compound is combined with an in vitro or cell-free system comprising KPP under conditions suitable for KPP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of KPP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding KPP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding KPP may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding KPP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding KPP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted  
5 as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress KPP, e.g., by secreting KPP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

### THERAPEUTICS

10 Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of KPP and kinases and phosphatases. In addition, examples of tissues expressing KPP can be found in Table 6. Therefore, KPP appears to play a role in cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers. In the treatment of disorders associated with increased KPP  
15 expression or activity, it is desirable to decrease the expression or activity of KPP. In the treatment of disorders associated with decreased KPP expression or activity, it is desirable to increase the expression or activity of KPP.

Therefore, in one embodiment, KPP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP.

20 Examples of such disorders include, but are not limited to, a cardiovascular disease such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive  
25 heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung  
30 anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung

- abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in
- 5 collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; an immune system disorder such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia,
- 10 isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis,
- 15 autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis,
- 20 myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder
- 25 such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative
- 30 intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental

retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic

5 myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a disorder affecting growth and development such as actinic keratosis, arteriosclerosis,

10 atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic

15 syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase

20 deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such as Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM<sub>2</sub> gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatosis, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipid adrenal hyperplasia, minimal change disease,

25 lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis,

30 hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder,

ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, uterus, leukemias such as multiple myeloma, and lymphomas such as Hodgkin's disease.

In another embodiment, a vector capable of expressing KPP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified KPP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of KPP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP including, but not limited to, those listed above.

In a further embodiment, an antagonist of KPP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of KPP. Examples of such disorders include, but are not limited to, those cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers described above. In one aspect, an antibody which specifically binds KPP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express KPP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding KPP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of KPP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of KPP may be produced using methods which are generally known in the art. In particular, purified KPP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind KPP. Antibodies to KPP may also be generated using

methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use. Single chain antibodies (e.g., from camels or llamas) may be potent  
5 enzyme inhibitors and may have advantages in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with KPP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species,  
10 various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to KPP  
15 have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of KPP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric  
20 molecule may be produced.

Monoclonal antibodies to KPP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J.  
25 Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc.  
30 Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce KPP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be



generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) *Proc. Natl. Acad. Sci. USA* 88:10134-10137.)

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:3833-3837; Winter, G. et al. (1991) *Nature* 349:293-299.)

Antibody fragments which contain specific binding sites for KPP may also be generated. For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) *Science* 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between KPP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering KPP epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for KPP. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of KPP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple KPP epitopes, represents the average affinity, or avidity, of the antibodies for KPP. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular KPP epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the KPP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of KPP, preferably in active form, from the antibody (Catty, D. (1988) *Antibodies, Volume I: A Practical Approach*, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) *A Practical Guide to Monoclonal Antibodies*, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of KPP-antibody  
5 complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding KPP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene  
10 expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding KPP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding KPP. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

15 In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995)  
20 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et  
25 al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding KPP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-  
30 linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene

Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in KPP expression or regulation causes disease, the expression of KPP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in KPP are treated by constructing mammalian expression vectors encoding KPP and introducing these vectors by mechanical means into KPP-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of KPP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). KPP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding KPP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to KPP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding KPP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) *J. Virol.* 61:1647-1650; Bender, M.A. et al. (1987) *J. Virol.* 61:1639-1646; Adam, M.A. and A.D. Miller (1988) *J. Virol.* 62:3802-3806; Dull, T. et al. (1998) *J. Virol.* 72:8463-8471; Zufferey, R. et al. (1998) *J. Virol.* 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding KPP to cells which have one or more genetic abnormalities with respect to the expression of KPP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) *Transplantation* 27:263-268). Potentially useful adenoviral vectors are

described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) *Annu. Rev. Nutr.* 19:511-544 and Verma, I.M. and N. Somia (1997) *Nature* 18:389:239-242, both incorporated by reference herein.

5 In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding KPP to target cells which have one or more genetic abnormalities with respect to the expression of KPP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing KPP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with  
10 ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92  
15 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) *J. Virol.* 73:519-532 and Xu, H. et al. (1994) *Dev. Biol.* 163:152-161, hereby incorporated by reference. The manipulation of cloned  
20 herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to  
25 deliver polynucleotides encoding KPP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the  
30 overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for KPP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of KPP-coding RNAs and the synthesis of high levels of KPP in vector transduced cells. While alphavirus infection is

typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of KPP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding KPP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA

sequences encoding KPP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

5 RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine,  
10 and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding KPP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not  
15 limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased KPP  
20 expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding KPP may be therapeutically useful, and in the treatment of disorders associated with decreased KPP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding KPP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in  
25 altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a  
30 library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding KPP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding KPP are assayed by

any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding KPP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or  
5 more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) *Nucleic Acids Res.* 28:E15) or a human  
10 cell line such as HeLa cell (Clarke, M.L. et al. (2000) *Biochem. Biophys. Res. Commun.* 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

15 Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) *Nat.*  
20 *Biotechnol.* 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which  
25 generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of KPP, antibodies to KPP, and mimetics, agonists, antagonists, or inhibitors of KPP.

30 The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.



Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising KPP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, KPP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example KPP or fragments thereof, antibodies of KPP, and agonists, antagonists or inhibitors of KPP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity.

The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu\text{g}$  to 100,000  $\mu\text{g}$ , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind KPP may be used for the diagnosis of disorders characterized by expression of KPP, or in assays to monitor patients being treated with KPP or agonists, antagonists, or inhibitors of KPP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for KPP include methods which utilize the antibody and a label to detect KPP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring KPP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of KPP expression. Normal or standard values for KPP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to KPP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of KPP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding KPP may be used for

diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of KPP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of KPP, and to monitor regulation of KPP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding KPP or closely related molecules may be used to identify nucleic acid sequences which encode KPP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding KPP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the KPP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:16-30 or from genomic sequences including promoters, enhancers, and introns of the KPP gene.

Means for producing specific hybridization probes for DNAs encoding KPP include the cloning of polynucleotide sequences encoding KPP or KPP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding KPP may be used for the diagnosis of disorders associated with expression of KPP. Examples of such disorders include, but are not limited to, a cardiovascular disease such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease,

cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic

5 obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes,

10 idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; an immune system disorder such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's

15 syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma,

20 atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis,

25 hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial,

30 fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy,

retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial

5 insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders,

10 dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a disorder

15 affecting growth and development such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental

20 retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis,

25 carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such as Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM<sub>2</sub> gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatosis, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid

30 adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease,

hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, uterus, leukemias such as multiple myeloma, and lymphomas such as Hodgkin's disease. The polynucleotide sequences encoding KPP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered KPP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding KPP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding KPP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding KPP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of KPP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding KPP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the

patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding KPP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding KPP, or a fragment of a polynucleotide complementary to the polynucleotide encoding KPP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding KPP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding KPP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the

high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations. (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641.)

Methods which may also be used to quantify the expression of KPP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic



profile.

In another embodiment, KPP, fragments of KPP, or antibodies specific for KPP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

5 A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No.  
10 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The  
15 resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention  
20 may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000)  
25 Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested  
30 compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not

necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed  
5 gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be  
10 quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global  
15 pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is  
20 achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot  
25 is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass  
30 spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for KPP to quantify the levels of KPP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoz, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) *Proc. Natl. Acad. Sci. USA*

94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding KPP may be used to  
5 generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a  
10 chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop  
15 genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic  
20 map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding KPP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as  
25 linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely  
30 localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation,

inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, KPP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between KPP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with KPP, or fragments thereof, and washed. Bound KPP is then detected by methods well known in the art. Purified KPP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding KPP specifically compete with a test compound for binding KPP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with KPP.

In additional embodiments, the nucleotide sequences which encode KPP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, including U. S. Ser. No. 60/282,119, U. S. Ser. No. 60/283,588, U. S. Ser. No. 60/283,759, U. S. Ser. No. 60/285,589, U. S. Ser. No. 60/287,037, U. S. Ser. No. 60/287,036, U. S. Ser. No. 60/288,608, U. S. Ser. No. 60/289,909, U. S. Ser. No. 60/292,246, and U. S. Ser. No. 60/288,712, are expressly incorporated by reference herein.

## EXAMPLES

### I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

## II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an

AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

5 Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) *Anal. Biochem.* 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner  
10 (Labsystems Oy, Helsinki, Finland).

### III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler  
15 (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides  
20 were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, *supra*, unit 7.7). Some of the cDNA sequences were selected for extension using the  
25 techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public  
30 databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from Homo sapiens, Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); hidden Markov

model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) *Nucleic Acids Res.* 29:41-43); and HMM-based protein domain databases such as SMART (Schultz et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:5857-5864; Letunic, I. et al. (2002) *Nucleic Acids Res.* 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary

5 structures of gene families. See, for example, Eddy, S.R. (1996) *Curr. Opin. Struct. Biol.* 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA

10 assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide

15 sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software

20 Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of

25 Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a

30 match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID



NO:16-30. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

#### IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative kinases and phosphatases were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode kinases and phosphatases, the encoded polypeptides were analyzed by querying against PFAM models for kinases and phosphatases. Potential kinases and phosphatases were also identified by homology to Incyte cDNA sequences that had been annotated as kinases and phosphatases. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

#### V. Assembly of Genomic Sequence Data with cDNA Sequence Data

##### "Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a

full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpi public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

#### "Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

#### **VI. Chromosomal Mapping of KPP Encoding Polynucleotides**

The sequences which were used to assemble SEQ ID NO:16-30 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:16-30 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for

Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

In this manner, SEQ ID NO:17 was mapped to chromosome 5 within the interval from 174.30 centiMorgans to qter, to chromosome 10 within the interval from 83.30 to 89.40 centiMorgans, and to chromosome 10 within the interval from 89.40 to 96.90 centiMorgans. More than one map location is reported for SEQ ID NO:17, indicating that sequences having different map locations were assembled into a single cluster. This situation occurs, for example, when sequences having strong similarity, but not complete identity, are assembled into a single cluster.

## VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum} \{ \text{length}(\text{Seq. 1}), \text{length}(\text{Seq. 2}) \}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding KPP are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding KPP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

#### VIII. Extension of KPP Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using

OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

- 5 Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ ,  
10 and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C,  
15 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100  $\mu$ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5  $\mu$ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar,  
20 Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5  $\mu$ l to 10  $\mu$ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

- 25 The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended  
30 clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-

well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

#### **IX. Identification of Single Nucleotide Polymorphisms in KPP Encoding Polynucleotides**

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:16-30 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezuelan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The

Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed  
5 no allelic variance in this population were not further tested in the other three populations.

#### X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:16-30 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide  
10 fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ - $^{32}$ P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech).  
15 An aliquot containing  $10^7$  counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16  
20 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

#### XI. Microarrays

25 The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), *supra*). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a  
30 procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science

270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

#### Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14  $\mu$ l 5X SSC/0.2% SDS.

#### Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are



amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5  $\mu$ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1  $\mu$ l of the array element DNA, at an average concentration of 100 ng/ $\mu$ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

#### Hybridization

Hybridization reactions contain 9  $\mu$ l of sample mixture consisting of 0.2  $\mu$ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140  $\mu$ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

#### Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a

resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate  
5 filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a  
10 cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed,  
15 the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a  
20 linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot  
25 is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

## **XII. Complementary Polynucleotides**

Sequences complementary to the KPP-encoding sequences, or any parts thereof, are used to  
30 detect, decrease, or inhibit expression of naturally occurring KPP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of KPP. To inhibit transcription, a

complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the KPP-encoding transcript.

### XIII. Expression of KPP

- 5 Expression and purification of KPP is achieved using bacterial or virus-based expression systems. For expression of KPP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element.
- 10 Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express KPP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of KPP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding KPP
- 15 by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA
- 20 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

- In most expression systems, KPP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized
- 25 glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from KPP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins
- 30 (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified KPP obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, XIX, XX, XXI, and XXII where applicable.

### XIV. Functional Assays

KPP function is assessed by expressing the sequences encoding KPP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of KPP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding KPP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding KPP and other genes of interest can be analyzed by northern analysis or microarray techniques.

#### **XV. Production of KPP Specific Antibodies**

KPP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the KPP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well  
5 described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the  
10 oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for anti-peptide and anti-KPP activity by, for example, binding the peptide or KPP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### XVI. Purification of Naturally Occurring KPP Using Specific Antibodies

15 Naturally occurring or recombinant KPP is substantially purified by immunoaffinity chromatography using antibodies specific for KPP. An immunoaffinity column is constructed by covalently coupling anti-KPP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

20 Media containing KPP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of KPP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/KPP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and KPP is collected.

#### 25 XVII. Identification of Molecules Which Interact with KPP

KPP, or biologically active fragments thereof, are labeled with  $^{125}\text{I}$  Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled KPP, washed, and any wells with labeled KPP complex are assayed. Data obtained using different concentrations of  
30 KPP are used to calculate values for the number, affinity, and association of KPP with the candidate molecules.

Alternatively, molecules interacting with KPP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) *Nature* 340:245-246, or using commercially available kits

based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

KPP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S.

5 Patent No. 6,057,101).

#### XVIII. Demonstration of KPP Activity

Generally, protein kinase activity is measured by quantifying the phosphorylation of a protein substrate by KPP in the presence of [ $\gamma$ - $^{32}\text{P}$ ]ATP. KPP is incubated with the protein substrate,  $^{32}\text{P}$ -ATP, and an appropriate kinase buffer. The  $^{32}\text{P}$  incorporated into the substrate is separated from  
10 free  $^{32}\text{P}$ -ATP by electrophoresis and the incorporated  $^{32}\text{P}$  is counted using a radioisotope counter. The amount of incorporated  $^{32}\text{P}$  is proportional to the activity of KPP. A determination of the specific amino acid residue phosphorylated is made by phosphoamino acid analysis of the hydrolyzed protein.

In one alternative, protein kinase activity is measured by quantifying the transfer of gamma phosphate from adenosine triphosphate (ATP) to a serine, threonine or tyrosine residue in a protein  
15 substrate. The reaction occurs between a protein kinase sample with a biotinylated peptide substrate and gamma  $^{32}\text{P}$ -ATP. Following the reaction, free avidin in solution is added for binding to the biotinylated  $^{32}\text{P}$ -peptide product. The binding sample then undergoes a centrifugal ultrafiltration process with a membrane which will retain the product-avidin complex and allow passage of free gamma  $^{32}\text{P}$ -ATP. The reservoir of the centrifuged unit containing the  $^{32}\text{P}$ -peptide product as retentate  
20 is then counted in a scintillation counter. This procedure allows the assay of any type of protein kinase sample, depending on the peptide substrate and kinase reaction buffer selected. This assay is provided in kit form (ASUA, Affinity Ultrafiltration Separation Assay, Transbio Corporation, Baltimore MD, U.S. Patent No. 5,869,275). Suggested substrates and their respective enzymes include but are not limited to: Histone H1 (Sigma) and p34<sup>cdc2</sup> kinase, Annexin I, Angiotensin (Sigma) and EGF receptor  
25 kinase, Annexin II and *src* kinase, ERK1 & ERK2 substrates and MEK, and myelin basic protein and ERK (Pearson, J.D. et al. (1991) Methods Enzymol. 200:62-81).

In another alternative, protein kinase activity of KPP is demonstrated in an assay containing KPP, 50  $\mu\text{l}$  of kinase buffer, 1  $\mu\text{g}$  substrate, such as myelin basic protein (MBP) or synthetic peptide substrates, 1 mM DTT, 10  $\mu\text{g}$  ATP, and 0.5  $\mu\text{Ci}$  [ $\gamma$ - $^{32}\text{P}$ ]ATP. The reaction is incubated at 30°C for  
30 30 minutes and stopped by pipetting onto P81 paper. The unincorporated [ $\gamma$ - $^{32}\text{P}$ ]ATP is removed by washing and the incorporated radioactivity is measured using a scintillation counter. Alternatively, the reaction is stopped by heating to 100°C in the presence of SDS loading buffer and resolved on a 12% SDS polyacrylamide gel followed by autoradiography. Incorporated radioactivity is corrected for

reactions carried out in the absence of KPP or in the presence of the inactive kinase, K38A. The amount of incorporated  $^{32}\text{P}$  is proportional to the activity of KPP.

In yet another alternative, adenylate kinase or guanylate kinase activity of KPP may be measured by the incorporation of  $^{32}\text{P}$  from  $[\gamma\text{-}^{32}\text{P}]\text{ATP}$  into ADP or GDP using a gamma radioisotope counter. KPP, in a kinase buffer, is incubated together with the appropriate nucleotide  
5 mono-phosphate substrate (AMP or GMP) and  $^{32}\text{P}$ -labeled ATP as the phosphate donor. The reaction is incubated at  $37^\circ\text{C}$  and terminated by addition of trichloroacetic acid. The acid extract is neutralized and subjected to gel electrophoresis to separate the mono-, di-, and triphosphonucleotide fractions. The diphosphonucleotide fraction is excised and counted. The radioactivity recovered is  
10 proportional to the activity of KPP.

In yet another alternative, other assays for KPP include scintillation proximity assays (SPA), scintillation plate technology and filter binding assays. Useful substrates include recombinant proteins tagged with glutathione transferase, or synthetic peptide substrates tagged with biotin. Inhibitors of KPP activity, such as small organic molecules, proteins or peptides, may be identified by such assays.

15 In another alternative, phosphatase activity of KPP is measured by the hydrolysis of para-nitrophenyl phosphate (PNPP). KPP is incubated together with PNPP in HEPES buffer pH 7.5, in the presence of 0.1%  $\beta$ -mercaptoethanol at  $37^\circ\text{C}$  for 60 min. The reaction is stopped by the addition of 6 ml of 10 N NaOH (Diamond, R.H. et al. (1994) Mol. Cell. Biol. 14:3752-62). Alternatively, acid phosphatase activity of KPP is demonstrated by incubating KPP-containing extract with 100  $\mu\text{l}$  of 10  
20 mM PNPP in 0.1 M sodium citrate, pH 4.5, and 50  $\mu\text{l}$  of 40 mM NaCl at  $37^\circ\text{C}$  for 20 min. The reaction is stopped by the addition of 0.5 ml of 0.4 M glycine/NaOH, pH 10.4 (Saftig, P. et al. (1997) J. Biol. Chem. 272:18628-18635). The increase in light absorbance at 410 nm resulting from the hydrolysis of PNPP is measured using a spectrophotometer. The increase in light absorbance is proportional to the activity of KPP in the assay.

25 In the alternative, KPP activity is determined by measuring the amount of phosphate removed from a phosphorylated protein substrate. Reactions are performed with 2 or 4 nM KPP in a final volume of 30  $\mu\text{l}$  containing 60 mM Tris, pH 7.6, 1 mM EDTA, 1 mM EGTA, 0.1%  $\beta$ -mercaptoethanol and 10  $\mu\text{M}$  substrate,  $^{32}\text{P}$ -labeled on serine/threonine or tyrosine, as appropriate. Reactions are initiated with substrate and incubated at  $30^\circ\text{C}$  for 10-15 min. Reactions are quenched with 450  $\mu\text{l}$  of  
30 4% (w/v) activated charcoal in 0.6 M HCl, 90 mM  $\text{Na}_4\text{P}_2\text{O}_7$ , and 2 mM  $\text{NaH}_2\text{PO}_4$ , then centrifuged at  $12,000 \times g$  for 5 min. Acid-soluble  $^{32}\text{Pi}$  is quantified by liquid scintillation counting (Sinclair, C. et al. (1999) J. Biol. Chem. 274:23666-23672).

#### XIX. Kinase Binding Assay

Binding of KPP to a FLAG-CD44 cyt fusion protein can be determined by incubating KPP with anti-KPP-conjugated immunoaffinity beads followed by incubating portions of the beads (having 10-20 ng of protein) with 0.5 ml of a binding buffer (20 mM Tris-HCL (pH 7.4), 150 mM NaCl, 0.1% bovine serum albumin, and 0.05% Triton X-100) in the presence of <sup>125</sup>I-labeled FLAG-CD44cyt fusion protein (5,000 cpm/ng protein ) at 4 °C for 5 hours. Following binding, beads were washed thoroughly in the binding buffer and the bead-bound radioactivity measured in a scintillation counter (Bourguignon, L.Y.W. et al. (2001) J. Biol. Chem. 276:7327-7336). The amount of incorporated <sup>32</sup>P is proportional to the amount of bound KPP.

#### XX. Identification of KPP Inhibitors

Compounds to be tested are arrayed in the wells of a 384-well plate in varying concentrations along with an appropriate buffer and substrate, as described in the assays in Example XVII. KPP activity is measured for each well and the ability of each compound to inhibit KPP activity can be determined, as well as the dose-response kinetics. This assay could also be used to identify molecules which enhance KPP activity.

#### XXI. Identification of KPP Substrates

A KPP "substrate-trapping" assay takes advantage of the increased substrate affinity that may be conferred by certain mutations in the PTP signature sequence of protein tyrosine phosphatases. KPP bearing these mutations form a stable complex with their substrate; this complex may be isolated biochemically. Site-directed mutagenesis of invariant residues in the PTP signature sequence in a clone encoding the catalytic domain of KPP is performed using a method standard in the art or a commercial kit, such as the MUTA-GENE kit from BIO-RAD. For expression of KPP mutants in *Escherichia coli*, DNA fragments containing the mutation are exchanged with the corresponding wild-type sequence in an expression vector bearing the sequence encoding KPP or a glutathione S-transferase (GST)-KPP fusion protein. KPP mutants are expressed in *E. coli* and purified by chromatography.

The expression vector is transfected into COS1 or 293 cells via calcium phosphate-mediated transfection with 20 µg of CsCl-purified DNA per 10-cm dish of cells or 8 µg per 6-cm dish. Forty-eight hours after transfection, cells are stimulated with 100 ng/ml epidermal growth factor to increase tyrosine phosphorylation in cells, as the tyrosine kinase EGFR is abundant in COS cells. Cells are lysed in 50 mM Tris-HCl, pH 7.5/5 mM EDTA/150 mM NaCl/1% Triton X-100/5 mM iodoacetic acid/10 mM sodium phosphate/10 mM NaF/5 µg/ml leupeptin/5 µg/ml aprotinin/1 mM benzamidine (1 ml per 10-cm dish, 0.5 ml per 6-cm dish). KPP is immunoprecipitated from lysates with an appropriate antibody. GST-KPP fusion proteins are precipitated with glutathione-Sepharose, 4 µg of



mAb or 10  $\mu$ l of beads respectively per mg of cell lysate. Complexes can be visualized by PAGE or further purified to identify substrate molecules (Flint, A.J. et al. (1997) Proc. Natl. Acad. Sci. USA 94:1680-1685).

## XXII. Enhancement/Inhibition of Protein Kinase Activity

- 5        Agonists or antagonists of PKIN activation or inhibition may be tested using assays described in section XVIII. Agonists cause an increase in PKIN activity and antagonists cause a decrease in PKIN activity.

- Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention.
- 10    Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

| Incyte Project ID | Polypeptide<br>SEQ ID NO: | Incyte<br>Polypeptide ID | Polynucleotide<br>SEQ ID NO: | Incyte<br>Polynucleotide<br>ID |
|-------------------|---------------------------|--------------------------|------------------------------|--------------------------------|
| 2763993           | 1                         | 2763993CD1               | 16                           | 2763993CB1                     |
| 3684162           | 2                         | 3684162CD1               | 17                           | 3684162CB1                     |
| 3736769           | 3                         | 3736769CD1               | 18                           | 3736769CB1                     |
| 7474632           | 4                         | 7474632CD1               | 19                           | 7474632CB1                     |
| 7472696           | 5                         | 7472696CD1               | 20                           | 7472696CB1                     |
| 7472343           | 6                         | 7472343CD1               | 21                           | 7472343CB1                     |
| 7480783           | 7                         | 7480783CD1               | 22                           | 7480783CB1                     |
| 7477063           | 8                         | 7477063CD1               | 23                           | 7477063CB1                     |
| 7475394           | 9                         | 7475394CD1               | 24                           | 7475394CB1                     |
| 7482884           | 10                        | 7482884CD1               | 25                           | 7482884CB1                     |
| 7494121           | 11                        | 7494121CD1               | 26                           | 7494121CB1                     |
| 6793486           | 12                        | 6793486CD1               | 27                           | 6793486CB1                     |
| 7494178           | 13                        | 7494178CD1               | 28                           | 7494178CB1                     |
| 7096516           | 14                        | 7096516CD1               | 29                           | 7096516CB1                     |
| 7474666           | 15                        | 7474666CD1               | 30                           | 7474666CB1                     |

Table 2

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: or PROTEOME ID NO: | Probability Score | Annotation  |
|------------------------|-----------------------|-----------------------------------|-------------------|---|
| 1                      | 2763993CD1            | g15215576                         | 0.0               | BMP-2 inducible kinase [Mus musculus]<br>Kearns, A.E. et al. (2001) Cloning and Characterization of a Novel Protein Kinase That Impairs Osteoblast Differentiation in Vitro. J. Biol. Chem. 276:42213-42218                       |
| 2                      | 3684162CD1            | g23903                            | 2.4E-271          | [Homo sapiens] 63kDa protein kinase<br>Gonzalez, F.A. et al., (1992) FEBS Lett. 304:170-178   |
| 3                      | 3736769CD1            | g9664225                          | 3.7E-27           | [Homo sapiens] FUSED serine/threonine kinase<br>Murone, M., et al. (2000) Nat. Cell Biol. 2:310-312   |
| 4                      | 7474632CD1            | g1304387                          | 1.0E-14           | [Saccharomyces cerevisiae var. diastaticus] glucoamylase<br>Lambrechts, M.G., et al. (1996) Proc. Natl. Acad. Sci. U.S.A. 93:8419-8424  |
| 5                      | 7472696CD1            | g14571717                         | 1.0E-124          | myosin light chain kinase (MLCK) [Homo sapiens]<br>[Homo sapiens] protein kinase HIPK2  |
| 6                      | 7472343CD1            | g11907599                         | 3.5E-87           | Wang, Y. et al., (2001) Biochim. Biophys. Acta 1518:168-172   |
| 7                      | 7480783CD1            | g4884492                          | 0.0               | [Rattus norvegicus] SCOP<br>Shimizu, K. et al. (1999) SCOP, a novel gene product expressed in a circadian manner in rat suprachiasmatic nucleus. FEBS Lett. 458:363-369   |
| 8                      | 7477063CD1            | g4115429                          | 9.2E-48           | [Rattus norvegicus] serine/threonine protein kinase   |
| 9                      | 7475394CD1            | g12005724                         | 6.9E-250          | [Homo sapiens] mixed lineage kinase MLK1  |
| 10                     | 7482884CD1            | g8918488                          | 8.1E-33           | [Rattus norvegicus] adenylate kinase isozyme 1  |
| 11                     | 7494121CD1            | g17385401                         | 0.0               | TPP alpha lipid phosphatase [Homo sapiens]<br>Walker, S.M. et al. (2001) TPP: a novel phosphoinositide 3-phosphatase. Biochem. J. 360: 277-283  |
| 12                     | 6793486CD1            | g5757703                          | 0.0               | [Mus musculus] syntrophin-associated serine-threonine protein kinase<br>Lumeng, C. et al. (1999) Interactions between beta 2 syntrophin and a family of microtubule-associated serine/threonine kinases. Nat. Neurosci. 2:611-617 |

Table 2

| Polypeptide SEQ<br>ID NO: | Incyte<br>Polypeptide ID | GenBank ID NO:<br>or PROTEOME<br>ID NO: | Probability<br>Score | Annotation  |
|---------------------------|--------------------------|---|----------------------|---|
| 13                        | 7494178CD1               | g7768151                                | 4.4E-12              | [Fagus sylvatica] protein phosphatase 2C (PP2C)<br>Lorenzo, O. et al. (2001) Plant Physiol. 125:1949-1956   |
|                           | 7494178CD1               | g12850332                               | 6.0E-94              | Protein phosphatase 2C containing protein~data source: Pfam, source<br>key: PF00481, evidence: ISS~putative [Mus musculus]<br>Carninci, P. and Hayashizaki, Y. (1999) High-efficiency full-length cDNA cloning.<br>Meth. Enzymol. 303:19-44 |
| 14                        | 7096516CD1               | g1457961                                | 0.0                  | [Mus musculus] receptor tyrosine kinase<br>Lee, A.M. et al. (1996) DNA Cell Biol. 15, 817-825   |
| 15                        | 7474666CD1               | g6688201                                | 9.3E-261             | [Homo sapiens] AMP-activated protein kinase gamma 3 subunit<br>Cheung, P.C. et al. (2000) Biochem J. 346:659-669  |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites  | Potential Glycosylation Sites                       | Signature Sequences, Domains and Motifs  | Analytical Methods and Databases |
|------------|-----------------------|---------------------|--|---|--|----------------------------------|
| 1          | 2763993CD1            | 1161                | S5 S56 S118 S228<br>S567 S599 S650<br>S659 S728 S740<br>S742 S750 S753<br>S754 S791 S802<br>S804 S904 S945<br>S949 S979 S1003<br>S1029 S1031 S1064<br>T52 T151 T203<br>T225 T244 T245<br>T279 T342 T588<br>T636 T723 T777<br>T850 T944 T985<br>T1004 T1007<br>T1011 T1037 Y806 | N277 N329 N330<br>N597 N608 N614<br>N631 N703 N1044 | Eukaryotic protein kinase domain: V51-F314   | HMMER_PFAM                       |
|            |                       |                     |  |   | Transmembrane domain: M234-K258<br>N-terminus is cytosolic                             | TMAP                             |
|            |                       |                     |  |   | Tyrosine kinase catalytic domain signature PR00109:<br>H170-L188, A247-S269, Y286-I308 | BLIMPS_PRINTS                    |
|            |                       |                     |  |   | PROTEIN KINASE DOMAIN DM00004 P53974 23-288: L57-I308                                  | BLAST_DOMO                       |
|            |                       |                     |  |   | PROTEIN KINASE DOMAIN DM00004 P40494 23-287: L57-I308                                  | BLAST_DOMO                       |
|            |                       |                     |  |   | PROTEIN KINASE DOMAIN DM00004 P38080 36-309: G72-I308                                  | BLAST_DOMO                       |
|            |                       |                     |  |   | PROTEIN KINASE DOMAIN DM00004 P50528 43-286: E55-I308                                  | BLAST_DOMO                       |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites  | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs  | Analytical Methods and Databases |
|------------|-----------------------|---------------------|--|-------------------------------|--|----------------------------------|
| 1          |                       |                     |  |                               | Serine/Threonine protein kinases active-site signature: I176-L188  | MOTIFS                           |
| 2          | 3684162CD1            | 587                 | S36 S58 S158 S180 S196 S266 S316 S348 S354 S363 S401 S403 S405 S414 S436 S519 T120 T224 T551 Y179 Y424 | N205 N343                     | Eukaryotic protein kinase domain: F20-M312   | HMMER_PFAM                       |
|            |                       |                     |  |                               | Transmembrane domain: K208-F229<br>N-terminus is cytosolic   | TMAP                             |
|            |                       |                     |  |                               | Protein kinases signatures and profile: H125-H178  | PROFILESCAN                      |
|            |                       |                     |  |                               | Tyrosine kinase catalytic domain signature PR00109: Y139-I157, I210-A232   | BLIMPS_PRINTS                    |
|            |                       |                     |  |                               | EXTRACELLULAR SIGNAL REGULATED KINASE TRANSFERASE SERINE/THREONINE PROTEIN ATP-BINDING CELL CYCLE ERK3 MAP PD023511: M312-A494                                   | BLAST_PRODROM                    |
|            |                       |                     |  |                               | KINASE TRANSFERASE PROTEIN SERINE/THREONINE PROTEIN ATP-BINDING II PHOSPHORYLATION CASEIN ALPHA CHAIN PD002608: H181-S313  | BLAST_PRODROM                    |
|            |                       |                     |  |                               | KINASE PROTEIN TRANSFERASE ATP-BINDING SERINE/THREONINE PROTEIN PHOSPHORYLATION RECEPTOR TYROSINE PROTEIN PRECURSOR TRANSMEMBRANE PD000001: Y184-Y311, V100-R173 | BLAST_PRODROM                    |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites  | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs                              | Analytical Methods and Databases |
|------------|-----------------------|---------------------|--|-------------------------------|--|----------------------------------|
| 2          |                       |                     |  |                               | PROTEIN KINASE DOMAIN DM00004[P31152]21-302: V21-A303                | BLAST_DOMO                       |
|            |                       |                     |  |                               | PROTEIN KINASE DOMAIN DM00004[P27704]21-306: V21-A303                | BLAST_DOMO                       |
|            |                       |                     |  |                               | PROTEIN KINASE DOMAIN DM00004[A56352]21-306: V21-A303                | BLAST_DOMO                       |
|            |                       |                     |  |                               | do KINASE; SERINE; ATP; THREONINE; DM03966[P31152]357-534: Y357-Q536 | BLAST_DOMO                       |
|            |                       |                     |  |                               | Protein kinases ATP-binding region signature L26-K49                 | MOTIFS                           |
|            |                       |                     |  |                               | Serine/Threonine protein kinases active-site signature: V145-I157    | MOTIFS                           |
| 3          | 3736769CD1            | 1275                | S82 S123 S174<br>S214 S223 S244<br>S291 S292 S335<br>S348 S362 S376<br>S401 S470 S659<br>S663 S783 S794<br>S829 S909 S951<br>S1067 S1110 S1137<br>S1176 S1209 S1236<br>S1256 T26 T35<br>T60 T134 T186<br>T228 T270 T366<br>T371 T415 T475<br>T577 T587 T787<br>T889 T898 T927<br>T1208 T1228 | N355 N827                     | Eukaryotic protein kinase domain: F4-W280                            | HMMER_PFAM                       |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites   | Potential Glycosylation Sites    | Signature Sequences, Domains and Motifs   | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---|----------------------------------|---|----------------------------------|
| 3          |                       |                     |   |                                  | Transmembrane domain: S701-Q725, T736-R764, E902-R925, N1100-L1128<br>N-terminus is cytosolic | TMAP                             |
|            |                       |                     |   |                                  | Tyrosine kinase catalytic domain signature PR00109: H111-L129, S200-E222, A249-W271, V75-A88  | BLIMPS_PRINTS                    |
|            |                       |                     |   |                                  | PROTEIN KINASE DOMAIN DM00004[P54644]122-362: G181-L269                                       | BLAST_DOMO                       |
|            |                       |                     |   |                                  | PROTEIN KINASE DOMAIN DM00004[P28178]155-393: G181-L269                                       | BLAST_DOMO                       |
|            |                       |                     |   |                                  | PROTEIN KINASE DOMAIN DM00004[P53355]15-257: E8-L129  | BLAST_DOMO                       |
|            |                       |                     |   |                                  | PROTEIN KINASE DOMAIN DM00004[P34885]380-621: E168-L269                                       | BLAST_DOMO                       |
| 4          | 7474632CD1            | 1406                | S11 S29 S135 S181<br>S202 S206 S234<br>S255 S329 S416<br>S491 S524 S532<br>S598 S619 S706<br>S729 S731 S747<br>S786 S800 S826<br>S837 S844 S850<br>S877 S946 S1064<br>S1399 T147 T287<br>T407 T417 T769<br>T1012 T1099<br>T1342 | N103 N667 N727<br>N848 N944 N988 | Eukaryotic protein kinase domain: E1119-H1155, G1258-C1327                                    | HMMER_PFAM                       |
|            |                       |                     |   |                                  | Tyrosine protein kinases specific active-site signature: I1141-L1153                          | MOTIFS                           |



Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites                           | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs   | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---|-------------------------------|---|----------------------------------|
| 5          | 7472696CD1            | 463                 | S90 S172 S237<br>S412 S416 S454<br>T112 T159 T265<br>Y262 | N263                          | signal_cleavage: M1-A25   | SPSCAN                           |
|            |                       |                     |   |                               | Eukaryotic protein kinase domain: V171-L426   | HMMER_PFAM                       |
|            |                       |                     |   |                               | Transmembrane domain: V345-P367   | TMAP                             |
|            |                       |                     |   |                               | N-terminus is non-cytosolic   |                                  |
|            |                       |                     |   |                               | Tyrosine kinase catalytic domain signature PR00109: M245-I258, H282-C300, T349-D371   | BLIMPS_PRINTS                    |
|            |                       |                     |   |                               | Phosphorylase kinase family signature PR01049: D259-I280, S418-H429, R317-P332  | BLIMPS_PRINTS                    |
|            |                       |                     |   |                               | KINASE PROTEIN TRANSFERASE ATP-BINDING SERINE/THREONINE PROTEIN PHOSPHORYLATION RECEPTOR TYROSINE PROTEIN PRECURSOR TRANSMEMBRANE PD000001: P332-S427, E260-R317, Y169-D249 | BLAST_PRODOM                     |
|            |                       |                     |   |                               | PROTEIN KINASE DOMAIN DM00004 P07313 298-541: E175-A417   | BLAST_DOMO                       |
|            |                       |                     |   |                               | PROTEIN KINASE DOMAIN DM00004 JN0583 727-969: E175-S416   | BLAST_DOMO                       |
|            |                       |                     |   |                               | PROTEIN KINASE DOMAIN DM00004 S07571 5152-5396: E175-S416   | BLAST_DOMO                       |
|            |                       |                     |   |                               | PROTEIN KINASE DOMAIN DM00004 P53355 15-257: E175-S416  | BLAST_DOMO                       |
|            |                       |                     |   |                               | Protein kinases ATP-binding region signature: L177-K200   | MOTIFS                           |
|            |                       |                     |   |                               | Serine/Threonine protein kinases active-site signature: I288-C300   | MOTIFS                           |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites  | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs  | Analytical Methods and Databases |
|------------|-----------------------|---------------------|--|-------------------------------|--|----------------------------------|
| 6          | 7472343CD1            | 565                 | S6 S32 S265 S319 S354 S365 S368 S433 S469 S479 S495 S511 S521 S522 S562 T21 T33 T304 Y283 Y361 | N500 N550                     | Eukaryotic protein kinase domain: E322-V347  | HMMER_PFAM                       |
|            |                       |                     |  |                               | Transmembrane domain: H71-V88<br>N-terminus is non-cytosolic   | TMAP                             |
|            |                       |                     |  |                               | Tyrosine kinase catalytic domain signature PR00109: R126-L144, V196-N218   | BLIMPS_PRINTS                    |
|            |                       |                     |  |                               | KINASE TRANSFERASE PROTEIN<br>SERINE/THREONINE PROTEIN ATP-BINDING II<br>PHOSPHORYLATION CASEIN ALPHA CHAIN<br>PD002608: Y170-P234, V321-F346        | BLAST_PRODROM                    |
|            |                       |                     |  |                               | PROTEIN KINASE NUCLEAR<br>SERINE/THREONINE PROTEIN<br>HOMEODOMAIN INTERACTING HOMEBOX<br>DNA-BINDING SERINE/THREONINE F20B6.8<br>PD042899: L239-H360 | BLAST_PRODROM                    |
|            |                       |                     |  |                               | PROTEIN KINASE DOMAIN<br>DM00004[Q09815]519-804; D12-K274, S301-S337   | BLAST_DOMO                       |
|            |                       |                     |  |                               | PROTEIN KINASE DOMAIN DM00004[P14680]371-694; I13-Y283   | BLAST_DOMO                       |
|            |                       |                     |  |                               | PROTEIN KINASE DOMAIN<br>DM00004[Q09690]700-985; D12-P255, S319-P338   | BLAST_DOMO                       |
|            |                       |                     |  |                               | PROTEIN KINASE DOMAIN DM00004[P49657]101-409; I13-P338   | BLAST_DOMO                       |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites   | Potential Glycosylation Sites                                 | Signature Sequences, Domains and Motifs  | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---|---|--|----------------------------------|
| 6          |                       |                     |   |   | Protein kinases ATP-binding region signature: L17-K40  | MOTIFS                           |
|            |                       |                     |   |   | Serine/Threonine protein kinases active-site signature: I132-L144  | MOTIFS                           |
| 7          | 7480783CD1            | 1319                | S28 S91 S181 S244<br>S248 S256 S477<br>S485 S545 S717<br>S803 S842 S856<br>S988 S1081 S1115<br>S1145 S1159 S1237<br>S1297 T9 T123<br>T187 T228 T319<br>T355 T387 T438<br>T448 T501 T561<br>T610 T733 T760<br>T830 T970 T1056<br>T1092 T1262<br>T1269 Y989 | N301 N324 N375<br>N596 N670 N698<br>N844 N1137 N1157<br>N1238 | Leucine Rich Repeat: Q365-T387, Q710-P731, L499-A520, Q296-S318, R688-P709, T733-T759, S477-S498, R246-D269, M617-L640, N342-Q364, Q665-K687, K522-S545, P567-L590, H436-E459, T319-L341, H641-E664, N591-E613 | HMMER-PFAM                       |
|            |                       |                     |   |   | Protein phosphatase 2C: G868-V1022   | HMMER-PFAM                       |
|            |                       |                     |   |   | Leucine-rich repeat signature: PR00019: L568-L581, L663-L676   | BLIMPS-PRINTS                    |
|            |                       |                     |   |   | PROTEIN PHOSPHATASE 2C:<br>DM00377 P49606 1686-2009: K866-E1033  | BLAST-DOMO                       |
|            |                       |                     |   |   | PROTEIN PHOSPHATASE 2C:<br>DM00377 P40371 67-332: W780-L1029, S756-G817  | RI,AST-DOMO                      |
|            |                       |                     |   |   | Leucine zipper pattern: L571-L592  | MOTIFS                           |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites                         | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs  | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---|-------------------------------|--|----------------------------------|
| 8          | 7477063CD1            | 414                 | S93 S380 S389<br>S390 T62 T85<br>T165 T242 T315<br>T407 | N58                           | Signal Peptide:M9-A39, M9-G38  | HMMER                            |
|            |                       |                     |   |                               | Protein kinase domain: Y98-R349  | HMMER_PPFAM                      |
|            |                       |                     |   |                               | Transmembrane domain: T271-D298  | TMAP                             |
|            |                       |                     |   |                               | Protein kinases signatures and profile<br>protein_kinase_tyr: L194-T249                | PROFILESKAN                      |
|            |                       |                     |   |                               | Tyrosine kinase catalytic domain signature PR00109:<br>Q173-Q186, F208-V226, V278-A300 | BLIMPS_PRINTS                    |
|            |                       |                     |   |                               | YABT PROTEIN PD102837: L114-V226   | BLAST_PRODUM                     |
|            |                       |                     |   |                               | PROTEIN KINASE DOMAIN DM00004  | BLAST_DOMO                       |
|            |                       |                     |   |                               | P24723 356-597: I101-F288  |                                  |
|            |                       |                     |   |                               | JC1446 20-261: R102-R349   |                                  |
|            |                       |                     |   |                               | P09217 254-502: L100-R349  |                                  |
|            |                       |                     |   |                               | Q05513 246-494: L100-R349  |                                  |
|            |                       |                     |   |                               | Gram-positive cocci surface proteins 'anchoring'<br>hexapeptide L43-A48                | MOTIFS                           |
|            |                       |                     |   |                               | Protein kinases ATP-binding region signature L104-K127                                 | MOTIFS                           |
|            |                       |                     |   |                               | Serine/Threonine protein kinases active-site signature<br>L214-V226                    | MOTIFS                           |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites   | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs  | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---|-------------------------------|--|----------------------------------|
| 9          | 7475394CD1            | 1036                | S58 S75 S167 S416<br>S441 S455 S536<br>S564 S643 S688<br>S770 S808 S845<br>S882 S932 S937<br>S949 S952 S998<br>S1026 T284 T299<br>T368 T399 T560<br>T719 T783 T893<br>T1022 Y330 Y724 | N282 N538 N565                | SH3 domain: G41-C100   | HMMER_PFAM                       |
|            |                       |                     |   |                               | Protein kinase domain: L124-L398   | HMMER_PFAM                       |
|            |                       |                     |   |                               | Transmembrane domain: P343-T362 Q715-H743  | TMAP                             |
|            |                       |                     |   |                               | Receptor tyrosine kinase<br>BL00239: E171-A218, L237-I259, W296-R345,<br>L350-I394<br>BL00240: P116-A164, E295-V342, V342-I394<br>BL00790: Q144-C197, S303-W335, L361-M409 | BLIMPS_BLOCKS                    |
|            |                       |                     |   |                               | Protein kinases signatures and profile<br>protein_kinase_1yr.prf: L237-T299  | PROFILES CAN                     |
|            |                       |                     |   |                               | Tyrosine kinase catalytic domain signature PR00109:<br>L200-A213, E253-L271, G306-I316, S325-I347,<br>C369-F391  | BLIMPS_PRINTS                    |
|            |                       |                     |   |                               | SH3 domain signature PR00452: G41-A51, D55-Q70,<br>D77-Q86, R88-C100   | BLIMPS_PRINTS                    |
|            |                       |                     |   |                               | KINASE DOMAIN SH3 MIXED LINEAGE<br>SERINE/THREONINE WITH LEUCINE ZIPPER<br>PROLINE PD024997: I401-D819   | BLAST_PRODROM                    |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites  | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs  | Analytical Methods and Databases |
|------------|-----------------------|---------------------|--|-------------------------------|--|----------------------------------|
| 9          |                       |                     |  |                               | KINASE PROTEIN TRANSFERASE ATP-BINDING SERINE/THREONINE PROTEIN PHOSPHORYLATION RECEPTOR TYROSINE PROTEIN PRECURSOR TRANSMEMBRANE PD000001: L237-A312, A120-F202, W310-Y344, K360-E396 | BLAST_PRODUM                     |
|            |                       |                     |  |                               | PROTEIN KINASE DOMAIN DM00004  A53800 119-368: L126-F391  I38044 100-349: L126-F391  | BLAST_DOMO                       |
|            |                       |                     |  |                               | ZIPPER MOTIF LEUCINE DM08113  I38044 392-721: R433-P776, R433-E750  A53800 411-846: R433-P805, R433-G846, P919-D955  | BLAST_DOMO                       |
|            |                       |                     |  |                               | Protein kinases ATP-binding region signature I130-K151   | MOTIFS                           |
|            |                       |                     |  |                               | Serine/Threonine protein kinases active-site signature I259-L271   | MOTIFS                           |
| 10         | 7482884CD1            | 293                 | S25 S60 S66 S79<br>S93 S98 S146 S247<br>S265 T105 T115<br>T188 T263 Y226 |                               | Signal peptide: M1-G63   | SPSCAN                           |
|            |                       |                     |  |                               | Adenylate kinase: V54-L211   | HMMER-PFAM                       |
|            |                       |                     |  |                               | Adenylate kinase protein BL00113: L53-I69, S79-K122, P127-V141, Q178-E208  | BLIMPS-BLOCKS                    |
|            |                       |                     |  |                               | Adenylate kinase signature: P111-A163  | PROFILES SCAN                    |
|            |                       |                     |  |                               | Adenylate kinase signature PR00094: I132-E148, R180-F195, Q197-L211, L53-S66, G81-R95  | BLIMPS-PRINTS                    |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites                            | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs   | Analytical Methods and Databases |
|------------|-----------------------|---------------------|--|-------------------------------|---|----------------------------------|
| 10         |                       |                     |  |                               | Adenylate kinase, transferase, ATP-binding, transphosphorylase:PD000657: V54-R174, G179-L211                            | BLAST-PRODOM                     |
|            |                       |                     |  |                               | Adenylate kinase DM00290: P00570 1-131: R48-R174 P12115 1-130: K50-R174 P15700 14-141: I51-R174 JC4181 1-133: R48-R174  | BLAST-DOMO                       |
|            |                       |                     |  |                               | Adenylate kinase signature: I132-Q143   | MOTIFS                           |
| 11         | 749412ICD1            | 550                 | S64 S224 S254 S286 S379 S423 T231 T243 T361 T374 T459 T541 | N375 N450 N499                | Transmembrane Segments: H86-I114 E125-Q152 L157-R185 L194-R221 G340-Y368 N-terminus is noncytosolic                     | TMAP                             |
|            |                       |                     |  |                               | Tyrosine specific protein phosphatases signature and profiles tyr_phosphatase.prf: V316-E371                            | PROFILESCAN                      |
|            |                       |                     |  |                               | PROTEIN HYDROLASE PHOSPHATASE MULTIPLE ADVANCED CANCERS PHOSPHORYLATION TENSIN PROTEINTYROSINE PTEN PD007685: L218-S286 | BLAST_PRODOME                    |
|            |                       |                     |  |                               | PROTEIN PHOSPHORYLATION AUXILIN COAT REPEAT KIAA0473 CYCLIN G-ASSOCIATED KINASE TRANSFERASE PD025411: S286-V400         | BLAST_PRODOME                    |
|            |                       |                     |  |                               | Cell attachment sequence R425-D427  | MOTIFS                           |
|            |                       |                     |  |                               | Tyrosine specific protein phosphatases active site I336-M348  | MOTIFS                           |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites   | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs                                  | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---|-------------------------------|--|----------------------------------|
| 12         | 6793486CD1            | 1547                | S28 S32 S60 S64<br>S80 S81 S90 S113<br>S142 S169 S362<br>S668 S673 S679<br>S687 S708 S716<br>S722 S735 S736<br>S741 S773 S777<br>S782 S799 S849<br>S903 S938 S947<br>S973 S1058 S1068<br>S1081 S1088 S1092<br>S1093 S1119 S1140<br>S1143 S1184 S1193<br>S1263 S1283 S1287<br>S1333 S1402 S1446<br>S1477 T126 T144<br>T376 T427 T443<br>T591 T677 T693<br>T694 T760 T842<br>T856 T980 T1052<br>T1099 T1107<br>T1397 T1521<br>T1535 Y1149 | N11 N1056 N1115               | PDZ (membrane-associated) domain (Also known as DHR or GLGF): P967-F1054 | HMMER_PFAM                       |
|            |                       |                     |   |                               | Protein kinase domain: F374-F647   | HMMER_PFAM                       |
|            |                       |                     |   |                               | Protein kinase C terminal domain: R648-D676                              | HMMER_PFAM                       |



Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites                                    | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs   | Analytical Methods and Databases |
|------------|-----------------------|---------------------|--|-------------------------------|---|----------------------------------|
| 12         |                       |                     |  |                               | Transmembrane segments: E205-T233 D569-F587<br>N-terminus cytosolic   | TMAP                             |
|            |                       |                     |  |                               | Tyrosine kinase catalytic domain signature PR00109: M451-K464, Y487-I505, V568-D590   | BLIMPS_PRINTS                    |
|            |                       |                     |  |                               | Octicosa peptide repeat protein PF00564: F374-F428, F438-L488   | BLIMPS_PPFAM                     |
|            |                       |                     |  |                               | MICROTUBULE ASSOCIATED TESTIS SPECIFIC SERINE/THREONINE PROTEIN KINASE<br>MAST205<br>PD135564: S13-Y182<br>PD182663: P740-H1002<br>PD069998: T1050-Y1149<br>PD041650: K183-D373                                       | BLAST_PRODOM                     |
|            |                       |                     |  |                               | PROTEIN KINASE DOMAIN DM00004<br>A54602/455-712: T376-G634<br>S42867/75-498: I377-T528, H534-F675, D863-Q890<br>P43565/796-1240: I377-M524, D544-G634, S1117-S1235, P1080-P1128<br>P05986/1-397: N372-K520, V547-D691 | BLAST_DOMO                       |
|            |                       |                     |  |                               | Serine/Threonine protein kinases active-site signature I493-I505  | MOTIFS                           |
| 13         | 7494178CD1            | 505                 | S93 S222 S387<br>S448 S487 T97<br>T301 T322 T413<br>T435 T472 Y415 | N345                          | Protein phosphatase 2C: F157-R295, V379-A449  | HMMER_PPFAM                      |
|            |                       |                     |  |                               | Transmembrane segments: V256-I284<br>N-terminus is non-cytosolic  | TMAP                             |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites   | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs  | Analytical Methods and Databases                                 |
|------------|-----------------------|---------------------|---|-------------------------------|--|--|
| 13         |                       |                     |   |                               | Protein phosphatase 2C p BL01032: S487-I496, Y153-G162, L231-A248, R372-D385, D420-D432<br>C42C1.2 PROTEIN PD140721: V274-K370, R145-E355, R22-V128<br>PROTEIN PHOSPHATASE 2C MAGNESIUM HYDROLASE MANGANESE MULTIGENE FAMILY PP2C ISOFORM PD001101: G366-S448, Q217-G310, F151-L181<br>PROTEIN PHOSPHATASE 2C DM00377<br>P49596  1-295: S222-G310, R372-S448, Y153-E186<br>P49598 88-398: A371-L467, K224-G310, Y152-K182<br>P36993  1-304: R372-V443, S222-G310, Y153-I177<br>S39781  1-304: R372-V443, S222-G310, Y153-I177<br>Ephrin receptor ligand binding domain: Q34-C207 | BLIMPS_BLOCKS<br>BLAST_PRODROM<br>BLAST_PRODROM<br>BLAST_PRODROM |
| 14         | 7096516CD1            | 1036                | S199 S228 S239<br>S377 S427 S492<br>S527 S647 S854<br>S855 S871 S881<br>S888 S928 S953<br>S972 S997 T61<br>T107 T120 T146<br>T158 T210 T245<br>T267 T302 T438<br>T534 T572 T605<br>T611 T664 T700<br>T843 T942 T965<br>T988 T1021 Y490<br>Y612 Y794 | N343 N397 N410<br>N756        |  | HMMER_PFAM   |
|            |                       |                     |   |                               | SAM domain (Sterile alpha motif): E959-R1023   | HMMER_PFAM   |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs  | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---------------------------------|-------------------------------|--|----------------------------------|
| 14         |                       |                     |                                 |                               | Fibronectin type III domain: Q440-S527, P332-S425  | HMIMER_PFAM                      |
|            |                       |                     |                                 |                               | Protein kinase domain: I631-V927   | HMIMER_PFAM                      |
|            |                       |                     |                                 |                               | Transmembrane domain: V6-P24 E180-Y204 G546-R574 N691-R718 K765-G793   | TMAP                             |
|            |                       |                     |                                 |                               | N-terminus is non-cytosolic.   |                                  |
|            |                       |                     |                                 |                               | Receptor tyrosine kinase BL00239: G638-S647, E680-A727, L775-R797, A800-D825, E828-Y877, N882-I926BL00240: H669-N723, Q774-V811, P827-E874, E874-I926BL00790: Q34-N55, D64-P115, K168-I221, P246-E270, C276-P323, I342-I368, C379-S422, S458-K483, G504-T534, P601-G640, P653-A706, L759-M778, L779-A800, A801-P827, G835-W867, E868-G892, Y893-H941, F982-H1025 | BLIMPS_BLOCKS                    |
|            |                       |                     |                                 |                               | Protein kinases signatures and profile<br>protein_kinase_tyr.prf: Q774-P827  | PROFILESKAN                      |
|            |                       |                     |                                 |                               | Tyrosine kinase catalytic domain signature PRO0109: V751-R764, Y788-V806, I838-I848, S857-E879, C901-F923  | BLIMPS_PRINTS                    |
|            |                       |                     |                                 |                               | KINASE RECEPTOR PRECURSOR TYROSINE<br>PROTEIN EPHRIN TRANSFERASE ATP<br>BINDING PHOSPHORYLATION<br>TRANSMEMBRANE GLYCOPROTEIN<br>PD001495: L37-C207<br>PD001679: K529-I631<br>PD149648: P208-R280  | BLAST_PRODROM                    |
|            |                       |                     |                                 |                               | EPH FAMILY Y PROTEIN PD002683: P332-D441   | BLAST_PRODROM                    |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites   | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs  | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---|-------------------------------|--|----------------------------------|
| 14         |                       |                     |   |                               | RECEPTOR TYROSINE KINASE CLASS V<br>DM00501<br>P54758 35-381: V36-G383<br>P29320 31-376: L37-C382<br>P29318 30-375: L37-C382<br>PROTEIN KINASE DOMAIN DM00004 P54758 632<br>920: I633-K922<br>EGF-like domain signature 2 C263-C276<br>Protein kinases ATP-binding region signature I637-K663<br>Tyrosine protein kinases specific active-site signature Y794-V806<br>Receptor tyrosine kinase class V signature 1 F188-P208<br>Receptor tyrosine kinase class V signature 2 C250-E270 | BLAST_DOMO                       |
| 15         | 7474666CD1            | 489                 | S27 S44 S46 S65<br>S128 S129 S200<br>S232 S236 S288<br>S292 S381 S397<br>S416 S439 S469<br>T87 T112 T192<br>T208 T242 T271<br>T323 T347 T355<br>T370 Y194 | N33 N290                      | CBS domain: Q278-H332, T425-V478, I353-A406, M197-H251   | HMMER_PFAM                       |
|            |                       |                     |   |                               | Transmembrane domain: V204-A226  | TMAP                             |
|            |                       |                     |   |                               | PROTEIN KINASE REPEAT CBS DOMAIN 5' AMP ACTIVATED GAMMA1 SUBUNIT AMPK CHAIN PD010762: Q182-F281  | BLAST_PRODOM                     |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs   | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---------------------------------|-------------------------------|---|----------------------------------|
| 15         |                       |                     |                                 |                               | PROTEIN SIMILAR AMP ACTIVATED KINASE R53.7 PD156234: L284-P481                                  | BLAST_PRODUM                     |
|            |                       |                     |                                 |                               | GAMMA-1; CAT3; KINASE; ACTIVATED; DM07032<br>P54619 I-330: I183-S480<br>P12904 I-321: R186-L479 | BLAST_DOMO                       |

Table 4

| Polynucleotide<br>SEQ ID NO./<br>Incyte ID/ Sequence<br>Length | Sequence Fragments  |
|--|---|
| 16/2763993CB1/<br>4188   | 1-403, 298-883, 370-607, 370-868, 469-929, 535-1124, 552-800, 611-1116, 674-751, 750-884, 772-1067, 797-1067, 838-1077, 885-1163, 1021-1468, 1035-1231, 1066-1231, 1091-1382, 1114-1352, 1114-1374, 1151-1667, 1151-1789, 1167-1231, 1186-1430, 1186-1950, 1192-1418, 1232-1299, 1232-1457, 1232-1470, 1232-1532, 1232-1950, 1258-1527, 1260-1375, 1276-1934, 1277-1554, 1352-1798, 1352-1801, 1352-1889, 1352-1936, 1392-1927, 1431-1722, 1431-1724, 1433-1755, 1469-1724, 1508-1753, 1514-1869, 1521-1671, 1609-1682, 1609-1685, 1640-1931, 1640-1951, 1647-2298, 1758-1921, 1847-2063, 2063-2502, 2082-2324, 2139-2624, 2160-2617, 2169-2413, 2171-2595, 2197-2617, 2198-2617, 2227-2406, 2489-2758, 2520-2795, 2688-3095, 2804-3157, 2828-3287, 2856-3114, 2856-3342, 2881-3138, 2881-3491, 2941-3226, 3022-3649, 3060-3603, 3067-3342, 3144-3656, 3158-3593, 3211-3640, 3223-3656, 3263-3446, 3304-3742, 3332-3522, 3341-4188, 3379-3670, 3571-3806  |
| 17/3684162CB1/<br>4675   | 1-417, 1-453, 1-480, 1-543, 1-605, 1-620, 9-538, 11-446, 31-478, 34-358, 35-454, 37-551, 39-502, 132-401, 132-651, 148-684, 189-689, 206-689, 209-689, 396-855, 431-4648, 525-792, 525-910, 668-941, 668-973, 668-1086, 668-1213, 668-1289, 668-1372, 696-1339, 701-1320, 714-1237, 827-1350, 836-1457, 855-1382, 967-1626, 977-1278, 1016-1665, 1034-1568, 1146-1613, 1241-1824, 1254-1789, 1271-1637, 1273-1921, 1280-2017, 1295-1794, 1314-1893, 1321-1893, 1336-1929, 1341-1941, 1349-1794, 1357-1904, 1360-1957, 1394-1942, 1404-1930, 1405-2071, 1413-2096, 1426-2029, 1426-2055, 1460-1991, 1471-2053, 1477-2170, 1513-2126, 1566-2107, 1581-2255, 1586-2129, 1604-2285, 1605-2206, 1622-2154, 1634-1888, 1656-2149, 1690-2299, 1700-2333, 1707-2231, 1711-2120, 1802-2212, 1825-2346, 1871-2401, 1871-2450, 1873-2401, 1936-2471, 1938-2248, 1968-2475, 2026-2759, 2086-2761, 2102-2466, 2107-2761, 2124-2761, 2151-2741, 2188-2761, 2193-2761, 2206-2761, 2224-2759, 2224-2761, 2258-2761, |
|  | 2271-2729, 2275-2729, 2310-2442, 2316-2759, 2317-2729, 2320-2742, 2323-2732, 2324-2634, 2329-2759, 2353-2727, 2376-2759, 2377-2761, 2385-2760, 2409-2729, 2422-2729, 2446-2760, 2580-2761, 2581-2761, 2766-3050, 2766-3108, 2766-3214, 2766-3220, 2766-3224, 2766-3233, 2767-3246, 2793-3115, 2800-3045, 2824-3115, 2837-3093, 2855-3490, 2865-3134, 2929-3317, 2937-3499, 2944-3499, 2951-3635, 2983-3574, 3000-3289, 3010-3341, 3010-3624, 3021-3256, 3034-3609, 3062-3494, 3067-3342, 3076-3643, 3123-3403, 3192-3783, 3214-3466, 3230-3797, 3235-3460, 3235-3566, 3235-3677, 3235-3698, 3256-3521, 3256-3533, 3259-3640, 3280-3479, 3283-3567, 3287-3861, 3302-3622, 3303-3555, 3310-3601, 3338-3577, 3338-3590, 3338-3864, 3346-3616, 3357-3759, 3382-3976, 3384-3630, 3422-3708, 3426-3998, 3461-3738, 3478-3741, 3578-3820, 3609-3877, 3612-4053, 3612-4154,   |

Table 4

| Polynucleotide<br>SEQ ID NO./<br>Incye ID/ Sequence<br>Length | Sequence Fragments   |
|---|--|
|   | 3644-3934, 3721-3966, 3721-3989, 3723-3981, 3723-4232, 3726-4269, 3741-3958, 3754-4040, 3773-4284, 3774-4006, 3774-4007, 3774-4305, 3783-4056, 3814-4056, 3814-4074, 3825-4102, 3825-4613, 3836-4500, 3837-4168, 3890-4454, 3893-4171, 3897-4539, 3903-4190, 3911-4135, 3924-4187, 3924-4227, 3931-4590, 3980-4463, 3980-4607, 3984-4273, 3990-4532, 4000-4602, 4008-4668, 4014-4640, 4035-4640, 4037-4275, 4083-4614, 4087-4642, 4099-4644, 4125-4640, 4138-4650, 4155-4610, 4162-4446, 4189-4668, 4191-4651, 4192-4651, 4195-4440, 4200-4675, 4209-4670, 4211-4453, 4217-4670, 4218-4660, 4225-4646, 4228-4484, 4230-4648, 4239-4474, 4258-4646, 4265-4651, 4277-4670, 4292-4650, 4297-4557, 4300-4648, 4303-4655, 4314-4559, 4314-4576, 4317-4651, 4332-4670, 4347-4650, 4347-4668, 4349-4648, 4356-4650, 4371-4635, 4394-4651, 4394-4665, 4397-4663, 4425-4651   |
| 18/3736769CBI/<br>4407  | 1-273, 89-549, 89-716, 102-476, 112-631, 116-380, 117-560, 117-676, 118-781, 125-402, 150-439, 167-487, 172-475, 189-764, 247-475, 473-520, 497-532, 497-675, 628-782, 693-957, 693-1195, 693-1277, 799-1200, 1040-1624, 1282-1746, 1318-1422, 1365-1892, 1416-1458, 1557-1892, 1565-1892, 1567-1886, 1576-1858, 1590-1892, 1618-1892, 1619-2004, 1674-1892, 1748-1892, 1813-2087, 1893-2557, 1893-2870, 2022-2087, 2558-2870, 2619-3213, 2652-2881, 2652-3170, 2665-3222, 2913-3255, 2948-3018, 2962-3225, 3034-3323, 3063-3307, 3069-3516, 3115-3671, 3118-3674, 3130-3354, 3130-3516, 3130-3685, 3131-3358, 3329-3831, 3430-3690, 3436-3811, 3459-3707, 3465-3683, 3465-3977, 3495-3768, 3566-4198, 3609-4128, 3610-4182, 3622-4186, 3692-4194, 3758-4198, 3759-4061, 3759-4068, 3759-4116, 3759-4180, 3759-4187, 3759-4211, 3759-4278, 3759-4307, 3759-4316, 3759-4337, 3759-4353, 3759-4363, 3772-4386, 3773-4185, 3780-4198, 3799-4366, 3803-4198, |
|   | 3819-4396, 3834-4196, 3840-4316, 3841-4296, 3860-4362, 3878-4397, 3914-4000, 3915-4194, 3966-4407, 3983-4402, 3986-4391, 4009-4401   |

Table 4

| Polynucleotide<br>SEQ ID NO./<br>Incyte ID/ Sequence<br>Length | Sequence Fragments  |
|--|---|
| 19/7474632CB1/<br>4795   | 1-290, 6-254, 111-685, 311-764, 516-1199, 729-1315, 816-1568, 1497-2185, 1530-1824, 1530-2009, 1530-2063, 1530-2126, 1530-2129, 1530-2138, 1530-2164, 1530-2168, 1530-2170, 1530-2191, 1536-2190, 1648-2063, 1748-1825, 1796-2466, 1796-2476, 1865-2476, 1973-2158, 1981-2655, 1982-2482, 2006-2512, 2016-2606, 2019-2616, 2049-2627, 2086-2605, 2089-2650, 2104-2655, 2114-2655, 2125-2177, 2137-2616, 2145-2613, 2145-2632, 2164-2653, 2164-2654, 2176-2794, 2256-2616, 2265-2401, 2282-2869, 2282-2921, 2301-2426, 2361-2482, 2361-2627, 2467-2707, 2561-3240, 2665-3321, 2676-3279, 2689-3244, 2696-3321, 2754-3042, 2803-3089, 2811-3057, 2992-3233, 3072-3271, 3095-3397, 3101-3376, 3234-3482, 3234-3543, 3312-3543, 3349-3951, 3721-3969, 3757-3980, 3757-3993, 3757-4127, 3757-4327, 3759-4331, 3777-4044, 3803-4094, 3831-4229, 3862-4092, 3862-4115, 3862-4139, 3862-4365, 3862-4422, 3900-3965, 3946-4157, 3948-4208, 3969-4215, 3969-4523, 3973-4558, 4045-4320, 4080-4669, 4084-4374, 4105-4639, 4112-4349, 4130-4402, 4134-4633, 4146-4781, 4154-4385, 4179-4617, 4202-4707, 4217-4794, 4220-4666, 4241-4713, 4256-4538, 4264-4496, 4264-4680, 4264-4794, 4269-4795, 4283-4795, 4284-4538, 4284-4553, 4290-4526, 4311-4613, 4312-4795, 4361-4795   |
| 20/7472696CB1/<br>2386   | 1-1888, 920-1813, 925-1032, 929-1780, 935-1032, 974-1032, 1032-1534, 1032-1810, 1032-1888, 1037-1729, 1228-1969, 1348-1378, 1348-1408, 1348-1412, 1348-1417, 1348-1567, 1348-1575, 1348-1659, 1348-1662, 1348-1694, 1348-1704, 1348-1713, 1348-1718, 1348-1734, 1348-1740, 1348-1746, 1348-1809, 1348-1810, 1348-1820, 1348-1833, 1348-1842, 1348-1852, 1348-1857, 1348-1866, 1348-1871, 1348-1875, 1348-1892, 1348-1910, 1348-1913, 1348-1917, 1348-1922, 1348-1926, 1348-1929, 1348-1940, 1348-1941, 1348-1948, 1348-1950, 1348-1954, 1348-1958, 1348-1960, 1348-1964, 1348-1969, 1348-1972, 1348-1975, 1348-1977, 1348-1987, 1348-1993, 1348-2013, 1348-2128, 1348-2131, 1348-2146, 1349-1810, 1350-1810, 1351-1969, 1352-1810, 1352-1858, 1352-1977, 1352-2103, 1352-2119, 1352-2131, 1353-1809, 1357-1965, 1364-1969, 1377-1578, 1377-1780, 1377-1807, 1377-1809, 1379-1449, 1379-1809, 1420-2338, 1429-1810, 1433-1810, 1434-1810, 1450-2375, 1471-1810, 1471-2062, 1531-2358, 1539-2362, 1565-1810, 1596-1810, 1630-2360, 1632-1727, 1645-1810, 1646-2365, 1648-2365, 1653-1810, 1659-2368, 1678-1800, 1720-1810, 1761-1810, 1761-2365, 1761-2386, 1780-1810, 1780-2380, 1780-2381, 1780-2382, 1780-2384, 1788-1810, 1790-2368, 1798-2370, 1817-2365, 1858-2291, 1882-2386, 1896-2386, 1902-2368, 1902-2386, 1912-2370, 1916-2368, 1923-2365, 1927-2365, 1938-2367, 1949-2386, 1978-2386, 1983-2386, 2001-2384, 2023-2365, 2041-2365, 2130-2196, 2259-2365, 2266-2368, 2266-2370, 2266-2386, 2287-2353 |



Table 4

| Polynucleotide<br>SEQ ID NO./<br>Incyte ID/ Sequence<br>Length | Sequence Fragments   |
|--|--|
| 21/7472343CB1/<br>3269   | 1-438, 1-458, 1-480, 46-660, 70-619, 85-523, 116-497, 119-504, 186-976, 212-977, 251-977, 266-826, 282-833, 305-735, 305-752, 354-785, 359-785, 359-977, 364-800, 393-665, 393-688, 406-977, 415-978, 416-975, 416-976, 418-977, 477-977, 505-977, 562-975, 935-1959, 945-1269, 949-1269, 1757-1946, 1757-1952, 1757-1959, 1757-1967, 1757-2051, 1757-2061, 1757-2074, 1757-2085, 1757-2086, 1757-2088, 1757-2091, 1757-2120, 1757-2180, 1757-2266, 1757-2267, 1757-2897, 1759-2432, 1761-2468, 1868-2271, 1871-2393, 1899-2512, 1909-2286, 1952-2628, 1987-2603, 2077-2729, 2188-2884, 2214-2883, 2220-2517, 2231-2897, 2239-2897, 2243-2897, 2245-2897, 2256-2897, 2260-2897, 2264-2897, 2319-2897, 2396-2897, 2601-2641, 2644-2897, 2646-2869, 2699-2739, 2765-2920, 2765-3267, 2765-3269   |
| 22/7480783CB1/<br>4888   | 1-625, 358-726, 381-1151, 414-726, 447-726, 1035-1553, 1053-1690, 1066-1152, 1076-1152, 1150-1254, 1150-1303, 1150-1336, 1150-1544, 1228-1254, 1254-1443, 1440-1731, 1459-2146, 1459-2187, 1459-2209, 1459-2210, 1998-2663, 2000-2544, 2000-2548, 2000-2558, 2000-2601, 2000-2631, 2000-2632, 2000-2661, 2000-2662, 2000-2663, 2000-2702, 2000-2703, 2000-2711, 2000-2761, 2000-2790, 2000-2812, 2002-2788, 2002-2819, 2009-2105, 2014-2814, 2022-2210, 2144-2210, 2152-2663, 2177-2620, 2182-2620, 2191-2771, 2237-3083, 2308-2368, 2406-3222, 2418-2568, 2418-2594, 2418-2721, 2418-2790, 2418-2828, 2418-2834, 2458-3221, 2476-2834, 2539-2569, 2540-3086, 2556-3117, 2604-4888, 2621-2814, 2621-2815, 2621-2819, 2621-2821, 2622-2821, 2623-2821, 2637-2834, 2640-3454, 2660-2821, 2694-3554, 2740-3571, 2789-3576, 2824-3576, 2832-3576, 2833-3576, 2836-3576, 2845-3576, 2857-3576, 2869-3576, 2870-3576, 2876-3576, 2885-3576, 2914-3576, 2917-3576, 2939-3508, 2939-3531, 2980-3398, 3010-3576, 3013-3377, 3119-3573, 3125-3576, 3650-4000, 3864-3991, 4055-4260, 4055-4363, 4055-4385, 4055-4590, 4059-4678, 4193-4725, 4405-4725 |
| 23/7477063CB1/<br>2380   | 1-507, 388-861, 388-1507, 583-658, 583-861, 583-862, 583-1539, 658-1410, 659-861, 659-862, 659-1539, 861-2246, 862-1410, 862-1539, 1166-1410, 1477-2315, 1477-2365, 1477-2368, 1477-2372, 1477-2380, 1487-2347, 1490-2372  |

Table 4

| Polynucleotide<br>SEQ ID NO./<br>Incyte ID/ Sequence<br>Length | Sequence Fragments   |
|--|--|
| 24/7475394CB1/<br>3111   | 1-805, 703-876, 703-938, 703-949, 703-954, 703-958, 703-1011, 703-1077, 703-1079, 739-1140, 740-897, 740-1035, 740-1065, 740-1079, 740-1136, 740-1140, 741-1075, 741-1140, 742-1060, 742-1065, 743-938, 743-1072, 756-962, 759-857, 759-1031, 759-1065, 759-1070, 759-1076, 759-1077, 759-1078, 759-1079, 759-1140, 760-958, 776-1065, 923-1079, 974-1694, 975-1692, 986-1693, 986-1694, 988-1693, 992-1693, 1003-1694, 1017-1047, 1106-1692, 1439-2172, 1489-2218, 1502-2093, 1518-2318, 1555-2316, 1604-2074, 1604-2330, 1604-2430, 1655-2355, 1702-2355, 1758-2241, 1841-2355, 1883-2352, 1884-2355, 1954-2355, 1988-2355, 2183-2932, 2185-2926, 2352-2932, 2506-2705, 2506-2963, 2506-3007, 2705-3111  |
| 25/7482884CB1/<br>1372   | 1-601, 308-566, 472-522, 472-527, 472-529, 472-530, 472-533, 472-536, 472-537, 472-540, 472-541, 472-544, 472-558, 472-593, 474-541, 475-541, 476-541, 478-952, 478-982, 478-1198, 478-1235, 478-1238, 481-1106, 510-1227, 510-1240, 510-1260, 510-1304, 510-1329, 510-1355, 510-1371, 510-1372, 528-924, 554-1189   |
| 26/7494121CB1/<br>1704   | 1-714, 15-495, 15-506, 15-519, 15-522, 15-665, 15-669, 15-687, 15-708, 91-815, 115-224, 116-699, 117-224, 118-144, 161-711, 161-980, 172-198, 197-224, 284-495, 285-441, 285-469, 285-484, 285-486, 285-495, 285-498, 286-1704, 321-1156, 331-433, 331-486, 335-429, 335-430, 335-444, 335-461, 335-495, 336-495, 338-468, 338-491, 338-495, 338-498, 339-495, 341-495, 355-495, 356-495, 357-495, 361-495, 362-495, 368-1041, 371-495, 373-495, 381-495, 423-495, 432-495, 781-987, 847-987, 1081-1704, 1139-1330, 1407-1500  |
| 27/6793486CB1/<br>5563   | 1-591, 477-678, 477-758, 511-593, 511-682, 511-5151, 594-682, 594-758, 727-1235, 759-998, 784-1390, 784-1391, 789-1339, 820-1390, 837-1391, 838-998, 844-1410, 848-1391, 935-1424, 946-1487, 960-1588, 999-1074, 999-1284, 1067-1564, 1067-1583, 1067-1621, 1067-1654, 1067-1655, 1067-1680, 1067-1686, 1067-1687, 1067-1691, 1067-1693, 1067-1694, 1067-1695, 1073-1695, 1075-1284, 1075-1386, 1091-1692, 1093-1689, 1101-1695, 1118-1695, 1157-1695, 1160-1695, 1256-1695, 1285-1386, 1285-1519, 1384-1784, 1384-1975, 1384-1988, 1384-2029, 1384-2071, 1386-1910, 1386-2183, 1387-1519, 1387-1587, 1476-2297, 1486-2082, 1497-2084, 1520-1587, 1520-1667, 1577-2117, 1588-1667, 1588-1876, 1597-2217, 1606-2037, 1660-2157, 1663-2346, 1668-2015, 1737-2421, 1806-2395, 1833-2323, 1877-2015, 1877-2148, 1958-2539, 2014-2539, 2016-2148, 2095-2539, 2199-2733, 2199-2798, 2219-2802, 2298-2539, 2315-2539, 2339-2539, 2406-2539, 2417-2539, 2417-2649, 2436-2560, 2540-2649, 2540-2732, 2540-2802, 2650-2802, 2650-3076, 2803-3076, 2803-3283, 3009-3648, 3140-3745, 3170-3871, 3238-3704, 3284-3513, 3284-3636, 3307-3881, 3327-3815, 3514-3636, 3514-3773, 3637-3996, 3774-3996, 3774-4185, 3997-4185, 4186-5151, 4186-5154, 4459-5066, 4471-5193, 4543-5191, 4572-5174, 4607-5182, 4609-5149, 4813-5563 |

Table 4

| Polynucleotide<br>SEQ ID NO./<br>Incyte ID/ Sequence<br>Length | Sequence Fragments  |
|--|---|
| 28/7494178CB1/<br>1697   | 1-724, 232-491, 286-509, 336-778, 336-989, 342-620, 400-975, 419-648, 431-956, 473-1016, 561-801, 561-1144, 678-1506, 721-973, 873-1099, 876-1149, 913-1488, 944-1240, 964-1532, 971-1670, 1008-1261, 1008-1478, 1016-1592, 1022-1265, 1048-1689, 1183-1680, 1294-1540, 1376-1497, 1376-1679, 1376-1697, 1492-1697  |
| 29/7096516CB1/<br>3280   | 1-269, 1-270, 1-271, 1-272, 3-270, 77-270, 193-392, 207-323, 338-821, 338-1001, 434-933, 476-1095, 543-1358, 618-1303, 839-1001, 949-1493, 957-1493, 1088-1493, 1093-1493, 1159-2093, 1494-1664, 1494-2093, 1776-2019, 1822-2642, 1908-2305, 1908-2339, 1908-2395, 1908-2404, 1908-2425, 1908-2427, 1908-2472, 1908-2548, 1908-2563, 1908-2564, 1908-2574, 1911-2568, 1928-2424, 1957-2424, 1957-2476, 1957-2695, 2060-2695, 2278-2999, 2672-3015, 2880-3046, 2880-3210, 2906-3210, 3113-3280   |
| 30/7474666CB1/<br>2781   | 1-722, 1-726, 1-791, 1-792, 1-806, 5-670, 7-522, 7-528, 7-538, 9-510, 9-540, 15-598, 18-502, 18-541, 57-696, 58-677, 67-636, 210-879, 223-879, 232-879, 258-872, 287-879, 307-876, 354-876, 360-879, 367-879, 374-1197, 377-1197, 418-1197, 423-1197, 482-879, 1043-1298, 1077-1352, 1079-1579, 1079-1584, 1079-1663, 1079-1696, 1079-1700, 1079-1843, 1083-1710, 1092-1455, 1209-1794, 1264-1616, 1300-1718, 1424-1905, 1438-1966, 1465-1947, 1475-2061, 1477-2014, 1486-2122, 1506-1879, 1514-1971, 1517-1776, 1517-2041, 1517-2056, 1517-2080, 1517-2132, 1585-2179, 1587-2098, 1604-1997, 1614-2256, 1646-2332, 1651-2287, 1667-2339, 1681-2248, 1689-2239, 1704-2248, 1713-2079, 1768-2302, 1797-2300, 1824-2494, 1827-2302, 1877-2453, 1877-2466, 1881-2298, 1886-2302, 1896-2302, 1901-2302, 1920-2302, 1960-2506, 1999-2506, 2109-2476, 2163-2418, 2224-2781, 2234-2703 |

Table 5

| Polynucleotide SEQ<br>ID NO: | Incyte Project ID: | Representative Library |
|------------------------------|--------------------|------------------------|
| 16                           | 2763993CB1         | KIDNTUT13              |
| 17                           | 3684162CB1         | BRAUNOR01              |
| 18                           | 3736769CB1         | LUNGNOT12              |
| 19                           | 7474632CB1         | NEURDNV02              |
| 21                           | 7472343CB1         | HIPONON02              |
| 22                           | 7480783CB1         | PROSUNJ01              |
| 24                           | 7475394CB1         | LIVRTUE01              |
| 25                           | 7482884CB1         | BRAHTDR03              |
| 26                           | 7494121CB1         | PROSNOT14              |
| 27                           | 6793486CB1         | BRABDIR01              |
| 28                           | 7494178CB1         | UTRSNOT01              |
| 29                           | 7096516CB1         | BRACDIR02              |
| 30                           | 7474666CB1         | BONRFET01              |

Table 6

| Library   | Vector   | Library Description   |
|-----------|----------|---|
| BONRFET01 | pINCY    | Library was constructed using RNA isolated from rib bone tissue removed from a Caucasian male fetus, who died from Patau's syndrome (trisomy 13) at 20-weeks' gestation.  |
| BRABDIR01 | pINCY    | Library was constructed using RNA isolated from diseased cerebellum tissue removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident. Patient history included Huntington's disease, emphysema, and tobacco abuse.   |
| BRACDIR02 | PCDNA2.1 | This random primed library was constructed using RNA isolated from diseased corpus callosum tissue removed from a 57-year-old Caucasian male who died from a cerebrovascular accident. Patient history included Huntington's disease and emphysema.   |
| BRAHTDR03 | PCDNA2.1 | This random primed library was constructed using RNA isolated from archaecortex, anterior hippocampus tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.   |
| BRAUNOR01 | pINCY    | This random primed library was constructed using RNA isolated from striatum, globus pallidus and posterior putamen tissue removed from an 81-year-old Caucasian female who died from a hemorrhage and ruptured thoracic aorta due to atherosclerosis. Pathology indicated moderate atherosclerosis involving the internal carotids, bilaterally; microscopic infarcts of the frontal cortex and hippocampus; and scattered diffuse amyloid plaques and neurofibrillary tangles, consistent with age. Grossly, the leptomeninges showed only mild thickening and hyalinization along the superior sagittal sinus. The remainder of the leptomeninges was thin and contained some congested blood vessels. Mild atrophy was found mostly in the frontal poles and lobes, and temporal lobes, bilaterally. Microscopically, there were pairs of Alzheimer type II astrocytes within the deep layers of the neocortex. There was increased satellitosis around neurons in the deep gray matter in the middle frontal cortex. The amygdala contained rare diffuse plaques and neurofibrillary tangles. |
|           |          | The posterior hippocampus contained a microscopic area of cystic cavitation with hemosiderin-laden macrophages surrounded by reactive gliosis. Patient history included sepsis, cholangitis, post-operative atelectasis, pneumonia CAD, cardiomegaly due to left ventricular hypertrophy, splenomegaly, arteriolonephrosclerosis, nodular colloidal goiter, emphysema, CHF, hypothyroidism, and peripheral vascular disease.  |

Table 6

| Library   | Vector      | Library Description  |
|-----------|-------------|--|
| HIPONON02 | PSPORT1     | This normalized hippocampus library was constructed from 1.13M independent clones from a hippocampus tissue library. RNA was isolated from the hippocampus tissue of a 72-year-old Caucasian female who died from an intracranial bleed. Patient history included nose cancer, hypertension, and arthritis. The normalization and hybridization conditions were adapted from Soares et al. (PNAS (1994) 91:9228).  |
| KIDNTUT13 | pINCY       | Library was constructed using RNA isolated from kidney tumor tissue removed from a 51-year-old Caucasian female during a nephroureterectomy. Pathology indicated a grade 3 renal cell carcinoma. Patient history included depressive disorder, hypoglycemia, and uterine endometriosis. Family history included calculus of the kidney, colon cancer, and type II diabetes.  |
| LIVRTUE01 | PCDNA2.1    | This 5' biased random primed library was constructed using RNA isolated from liver tumor tissue removed from a 72-year-old Caucasian male during partial hepatectomy. Pathology indicated metastatic grade 2 (of 4) neuroendocrine carcinoma forming a mass. The patient presented with metastatic liver cancer. Patient history included benign hypertension, type I diabetes, prostatic hyperplasia, prostate cancer, alcohol abuse in remission, and tobacco abuse in remission. Previous surgeries included destruction of a pancreatic lesion, closed prostatic biopsy, transurethral prostatectomy, removal of bilateral testes and total splenectomy. Patient medications included Eulexin, Hytrin, Proscar, Ecotrin, and insulin. Family history included atherosclerotic coronary artery disease and acute myocardial infarction in the mother; atherosclerotic coronary artery disease and type II diabetes in the father. |
| LUNGNOT12 | pINCY       | Library was constructed using RNA isolated from lung tissue removed from a 78-year-old Caucasian male during a segmental lung resection and regional lymph node resection. Pathology indicated fibrosis pleura was puckered, but not invaded. Pathology for the associated tumor tissue indicated an invasive pulmonary grade 3 adenocarcinoma. Patient history included cerebrovascular disease, arteriosclerotic coronary artery disease, thrombophlebitis, chronic obstructive pulmonary disease, and asthma. Family history included intracranial hematoma, cerebrovascular disease, arteriosclerotic coronary artery disease, and type I diabetes.  |
| NEURDNV02 | PCR2-TOPOTA | Library was constructed using pooled cDNA from different donors. cDNA was generated using mRNA isolated from pooled skeletal muscle tissue removed from ten 21 to 57-year-old Caucasian male and female donors who died from sudden death; from pooled thymus tissue removed from nine 18 to 32-year-old Caucasian male and female donors who died from sudden death; from pooled liver tissue removed from 32 Caucasian male and female fetuses who died at 18-24 weeks gestation due to spontaneous abortion; from kidney tissue removed from 59 Caucasian male and female fetuses who died at 20-33 weeks gestation due to spontaneous abortion; and from brain tissue removed from a Caucasian male fetus who died at 23 weeks gestation due to fetal demise.  |

Table 6

| Library   | Vector  | Library Description   |
|-----------|---------|---|
| PROSNOT14 | pINCY   | Library was constructed using RNA isolated from diseased prostate tissue removed from a 60-year-old Caucasian male during radical prostatectomy and regional lymph node excision. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 3+4). The patient presented with elevated prostate specific antigen (PSA). Patient history included a kidney cyst and hematuria. Family history included benign hypertension, cerebrovascular disease, and arteriosclerotic coronary artery disease. |
| PROSUNJ01 | PIGEN   | This random primed 5' cap isolated library was constructed using RNA isolated from an untreated LNCaP cell line, derived from prostate carcinoma with metastasis to the left supraclavicular lymph nodes, removed from a 50-year-old Caucasian male (Schering).   |
| UTRSNOT01 | PSPORT1 | Library was constructed using RNA isolated from the uterine tissue of a 59-year-old female who died of a myocardial infarction. Patient history included cardiomyopathy, coronary artery disease, previous myocardial infarctions, hypercholesterolemia, hypotension, and arthritis.  |

Table 7

| Program           | Description   | Reference  | Parameter Threshold   |
|-------------------|---|--|---|
| ABI FACTURA       | A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.  | Applied Biosystems, Foster City, CA.   |   |
| ABI/PARACEL FDF   | A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.   | Applied Biosystems, Foster City, CA;<br>Paracel Inc., Pasadena, CA.  | Mismatch <50%   |
| ABI AutoAssembler | A program that assembles nucleic acid sequences.  | Applied Biosystems, Foster City, CA.   |   |
| BLAST             | A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.                    | Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.  | ESTs: Probability value=1.0E-8 or less<br>Full Length sequences: Probability value= 1.0E-10 or less   |
| FASTA             | A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch. | Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.              | ESTs: fasta E value=1.06E-6<br>Assembled ESTs: fasta Identity= 95% or greater and<br>Match length=200 bases or greater;<br>fastx E value=1.0E-8 or less<br>Full Length sequences:<br>fastx score=100 or greater |
| BLIMPS            | A BLOCKS IMPROVED Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.            | Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424. | Probability value=1.0E-3 or less  |
| HMMER             | An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART, and TIGRFAM.  | Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.   | PFAM, INCY, SMART, or TIGRFAM<br>hits: Probability value=1.0E-3 or less<br>Signal peptide hits: Score= 0 or greater   |



Table 7 (cont.)

| Program     | Description   | Reference  | Parameter Threshold   |
|-------------|---|--|---|
| ProfileScan | An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.   | Gribskov, M. et al. (1988) CABIOS 4:61-66;<br>Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.                                    | Normalized quality score > GCG-specified "HIGH" value for that particular Prosite motif.<br>Generally, score=1.4-2.1. |
| Phred       | A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.  | Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.  |   |
| Phrap       | A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences. | Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.             | Score=120 or greater;<br>Match length=56 or greater   |
| Consed      | A graphical tool for viewing and editing Phrap assemblies.  | Gordon, D. et al. (1998) Genome Res. 8:195-202.  |   |
| SPScan      | A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.  | Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.  | Score=3.5 or greater  |
| TMAP        | A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.   | Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.   |   |
| TMHMMER     | A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.   | Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182. |   |
| Motifs      | A program that searches amino acid sequences for patterns that matched those defined in Prosite.  | Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.   |   |

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
  - a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15,
  - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10 and SEQ ID NO:12-13,
  - c) a polypeptide comprising a naturally occurring amino acid sequence at least 91.5% identical to the amino acid sequence consisting of SEQ ID NO:11,
  - d) a polypeptide comprising a naturally occurring amino acid sequence at least 97% identical to the amino acid sequence consisting of SEQ ID NO:14,
  - e) a polypeptide comprising a naturally occurring amino acid sequence at least 99% identical to the amino acid sequence consisting of SEQ ID NO:15,
  - f) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and
  - g) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.
2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.
3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
9. A method of producing a polypeptide of claim 1, the method comprising:
- 5       a)       culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
- b)       recovering the polypeptide so expressed.
- 10       10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.
11. An isolated antibody which specifically binds to a polypeptide of claim 1.
- 15       12. An isolated polynucleotide selected from the group consisting of:
- a)       a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30,
- b)       a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ
- 20       ID NO:16-25 and SEQ ID NO:27-30,
- c)       a polynucleotide comprising a naturally occurring polynucleotide sequence at least 91.5% identical to the polynucleotide sequence of SEQ ID NO:26,
- d)       a polynucleotide complementary to a polynucleotide of a),
- e)       a polynucleotide complementary to a polynucleotide of b), and
- 25       f)       an RNA equivalent of a)-d).
13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.
- 30       14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:
- a)       hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and

which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

- b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15 15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide  
10 having a sequence of a polynucleotide of claim 12, the method comprising:

- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and  
b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

15

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence  
20 selected from the group consisting of SEQ ID NO:1-15.

19. A method for treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment the composition of claim 17.

25

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and  
b) detecting agonist activity in the sample.

30

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment a composition of claim 21.

5           23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting antagonist activity in the sample.

10           24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional KPP, comprising administering to a patient in need of such treatment a composition of claim 24.

15

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- 20           b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

- 25           a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test
- 30           compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

- a) treating a biological sample containing nucleic acids with the test compound,
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,
- c) quantifying the amount of hybridization complex, and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of KPP in a biological sample, the method comprising:

- a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and
- b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

- a) a chimeric antibody,
- b) a single chain antibody,

- c) a Fab fragment,
- d) a F(ab')<sub>2</sub> fragment, or
- e) a humanized antibody.

5           32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of KPP in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

10           34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of KPP in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

15           36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- 20       b) isolating antibodies from said animal, and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.

25           37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

30           39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

- b) isolating antibody producing cells from the animal,
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,
- d) culturing the hybridoma cells, and
- 5 e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.

10 40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

15 43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from  
20 the group consisting of SEQ ID NO:1-15 in a sample, the method comprising:

- a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of  
25 SEQ ID NO:1-15 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15 from a sample, the method comprising:

- a) incubating the antibody of claim 11 with a sample under conditions to allow specific  
30 binding of the antibody and the polypeptide, and
- b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.



46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

- a) labeling the polynucleotides of the sample,
- b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and
- c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

10

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

15

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

20

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

25

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

30

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.
71. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:16.
- 5 72. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:17.
73. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:18.
74. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:19.
- 10 75. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:20.
76. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:21.
- 15 77. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:22.
78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:23.
79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.
- 20 80. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:25.
81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:26.
- 25 82. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:27.
83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:28.
84. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:29.
- 30 85. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:30.

## &lt;110&gt; INCYTE GENOMICS, INC.

Recipon, Shirley A.  
Burrill, John D.  
Marcus, Gregory A.  
Zingler, Kurt A.  
Tang, Y. Tom  
Thornton, Michael  
Borowsky, Mark L.  
Baughn, Mariah R.  
Burford, Neil  
Lee, Soo Yeun  
Bandman, Olga  
Hafalia, April J.A.  
Yao, Monique G.  
Ramkumar, Jayalaxmi  
Walia, Narinder K.  
Lu, Dyung Aina M.  
Arvizu, Chandra S.  
Ison, Craig H.  
Ding, Li  
Lu, Yan  
Gururajan, Rajagopal  
Walsh, Roderick T.  
Gandhi, Ameena R.  
Swarnakar, Anita  
Forsythe, Ian J.  
Yue, Henry  
Au-Young, Janice K.  
Elliott, Vicki S.  
Lee, Sally

## &lt;120&gt; KINASES AND PHOSPHATASES

&lt;130&gt; PI-0398 PCT

&lt;140&gt; To Be Assigned

&lt;141&gt; Herewith

<150> 60/282,119; 60/283,588; 60/283,759; 60/285,589; 60/287,037;  
60/287,036; 60/288,608; 60/288,712; 60/289,909; 60/292,246<151> 2001-04-06; 2001-04-13; 2001-04-13; 2001-04-20; 2001-04-27;  
2001-04-27; 2001-05-04; 2001-05-04; 2001-05-09; 2001-05-17

&lt;160&gt; 30

&lt;170&gt; PERL Program

&lt;210&gt; 1

&lt;211&gt; 1161

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2763993CD1

&lt;400&gt; 1

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |    |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|
| Met | Lys | Lys | Phe | Ser | Arg | Met | Pro | Lys | Ser | Glu | Gly | Gly | Ser | Gly | 1   | 5   | 10  | 15 |
| Gly | Gly | Ala | Ala | Gly | Gly | Gly | Ala | Gly | Gly | Ala | Gly | Ala | Gly | Ala | 20  | 25  | 30  |    |
| Gly | Cys | Gly | Ser | Gly | Gly | Ser | Ser | Val | Gly | Val | Arg | Val | Phe | Ala | 35  | 40  | 45  |    |
| Val | Gly | Arg | His | Gln | Val | Thr | Leu | Glu | Glu | Ser | Leu | Ala | Glu | Gly | 50  | 55  | 60  |    |
| Gly | Phe | Ser | Thr | Val | Phe | Leu | Val | Arg | Thr | His | Gly | Gly | Ile | Arg | 65  | 70  | 75  |    |
| Cys | Ala | Leu | Lys | Arg | Met | Tyr | Val | Asn | Asn | Met | Pro | Asp | Leu | Asn | 80  | 85  | 90  |    |
| Val | Cys | Lys | Arg | Glu | Ile | Thr | Ile | Met | Lys | Glu | Leu | Ser | Gly | His | 95  | 100 | 105 |    |
| Lys | Asn | Ile | Val | Gly | Tyr | Leu | Asp | Cys | Ala | Val | Asn | Ser | Ile | Ser | 110 | 115 | 120 |    |
| Asp | Asn | Val | Trp | Glu | Val | Leu | Ile | Leu | Met | Glu | Tyr | Cys | Arg | Ala | 125 | 130 | 135 |    |
| Gly | Gln | Val | Val | Asn | Gln | Met | Asn | Lys | Lys | Leu | Gln | Thr | Gly | Phe | 140 | 145 | 150 |    |
| Thr | Glu | Pro | Glu | Val | Leu | Gln | Ile | Phe | Cys | Asp | Thr | Cys | Glu | Ala | 155 | 160 | 165 |    |
| Val | Ala | Arg | Leu | His | Gln | Cys | Lys | Thr | Pro | Ile | Ile | His | Arg | Asp | 170 | 175 | 180 |    |
| Leu | Lys | Val | Glu | Asn | Ile | Leu | Leu | Asn | Asp | Gly | Gly | Asn | Tyr | Val | 185 | 190 | 195 |    |
| Leu | Cys | Asp | Phe | Gly | Ser | Ala | Thr | Asn | Lys | Phe | Leu | Asn | Pro | Gln | 200 | 205 | 210 |    |
| Lys | Asp | Gly | Val | Asn | Val | Val | Glu | Glu | Glu | Ile | Lys | Lys | Tyr | Thr | 215 | 220 | 225 |    |
| Thr | Leu | Ser | Tyr | Arg | Ala | Pro | Glu | Met | Ile | Asn | Leu | Tyr | Gly | Gly | 230 | 235 | 240 |    |
| Lys | Pro | Ile | Thr | Thr | Lys | Ala | Asp | Ile | Trp | Ala | Leu | Gly | Cys | Leu | 245 | 250 | 255 |    |
| Leu | Tyr | Lys | Leu | Cys | Phe | Phe | Thr | Leu | Pro | Phe | Gly | Glu | Ser | Gln | 260 | 265 | 270 |    |
| Val | Ala | Ile | Cys | Asp | Gly | Asn | Phe | Thr | Ile | Pro | Asp | Asn | Ser | Arg | 275 | 280 | 285 |    |
| Tyr | Ser | Arg | Asn | Ile | His | Cys | Leu | Ile | Arg | Phe | Met | Leu | Glu | Pro | 290 | 295 | 300 |    |
| Asp | Pro | Glu | His | Arg | Pro | Asp | Ile | Phe | Gln | Val | Ser | Tyr | Phe | Ala | 305 | 310 | 315 |    |
| Phe | Lys | Phe | Ala | Lys | Lys | Asp | Cys | Pro | Val | Ser | Asn | Ile | Asn | Asn | 320 | 325 | 330 |    |
| Ser | Ser | Ile | Pro | Ser | Ala | Leu | Pro | Glu | Pro | Met | Thr | Ala | Ser | Glu | 335 | 340 | 345 |    |
| Ala | Ala | Ala | Arg | Lys | Ser | Gln | Ile | Lys | Ala | Arg | Ile | Thr | Asp | Thr | 350 | 355 | 360 |    |
| Ile | Gly | Pro | Thr | Glu | Thr | Ser | Ile | Ala | Pro | Arg | Gln | Arg | Pro | Lys | 365 | 370 | 375 |    |
| Ala | Asn | Ser | Ala | Thr | Thr | Ala | Thr | Pro | Ser | Val | Leu | Thr | Ile | Gln | 380 | 385 | 390 |    |
| Ser | Ser | Ala | Thr | Pro | Val | Lys | Val | Leu | Ala | Pro | Gly | Glu | Phe | Gly | 395 | 400 | 405 |    |
| Asn | His | Arg | Pro | Lys | Gly | Ala | Leu | Arg | Pro | Gly | Asn | Gly | Pro | Glu | 410 | 415 | 420 |    |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ile | Leu | Leu | Gly | Gln | Gly | Pro | Pro | Gln | Gln | Pro | Pro | Gln | Gln | His |
|     |     |     |     | 425 |     |     |     |     |     | 430 |     |     |     | 435 |
| Arg | Val | Leu | Gln | Gln | Leu | Gln | Gln | Gly | Asp | Trp | Arg | Leu | Gln | Gln |
|     |     |     |     | 440 |     |     |     |     |     | 445 |     |     |     | 450 |
| Leu | His | Leu | Gln | His | Arg | His | Pro | His | Gln | Gln | Gln | Gln | Gln | Gln |
|     |     |     |     | 455 |     |     |     |     |     | 460 |     |     |     | 465 |
| Gln | Gln | Gln | Gln | Gln | Gln | Gln | Gln | Gln | Gln | Gln | Gln | Gln | Gln | Gln |
|     |     |     |     | 470 |     |     |     |     |     | 475 |     |     |     | 480 |
| Gln | Gln | Gln | Gln | Gln | Gln | His | His | His | His | His | His | His | His | Leu |
|     |     |     |     | 485 |     |     |     |     |     | 490 |     |     |     | 495 |
| Leu | Gln | Asp | Ala | Tyr | Met | Gln | Gln | Tyr | Gln | His | Ala | Thr | Gln | Gln |
|     |     |     |     | 500 |     |     |     |     |     | 505 |     |     |     | 510 |
| Gln | Gln | Met | Leu | Gln | Gln | Gln | Phe | Leu | Met | His | Ser | Val | Tyr | Gln |
|     |     |     |     | 515 |     |     |     |     |     | 520 |     |     |     | 525 |
| Pro | Gln | Pro | Ser | Ala | Ser | Gln | Tyr | Pro | Thr | Met | Met | Pro | Gln | Tyr |
|     |     |     |     | 530 |     |     |     |     |     | 535 |     |     |     | 540 |
| Gln | Gln | Ala | Phe | Phe | Gln | Gln | Gln | Met | Leu | Ala | Gln | His | Gln | Pro |
|     |     |     |     | 545 |     |     |     |     |     | 550 |     |     |     | 555 |
| Ser | Gln | Gln | Gln | Ala | Ser | Pro | Glu | Tyr | Leu | Thr | Ser | Pro | Gln | Glu |
|     |     |     |     | 560 |     |     |     |     |     | 565 |     |     |     | 570 |
| Phe | Ser | Pro | Ala | Leu | Val | Ser | Tyr | Thr | Ser | Ser | Leu | Pro | Ala | Gln |
|     |     |     |     | 575 |     |     |     |     |     | 580 |     |     |     | 585 |
| Val | Gly | Thr | Ile | Met | Asp | Ser | Ser | Tyr | Ser | Ala | Asn | Arg | Ser | Val |
|     |     |     |     | 590 |     |     |     |     |     | 595 |     |     |     | 600 |
| Ala | Asp | Lys | Glu | Ala | Ile | Ala | Asn | Phe | Thr | Asn | Gln | Lys | Asn | Ile |
|     |     |     |     | 605 |     |     |     |     |     | 610 |     |     |     | 615 |
| Ser | Asn | Pro | Pro | Asp | Met | Ser | Gly | Trp | Asn | Pro | Phe | Gly | Glu | Asp |
|     |     |     |     | 620 |     |     |     |     |     | 625 |     |     |     | 630 |
| Asn | Phe | Ser | Lys | Leu | Thr | Glu | Glu | Glu | Leu | Leu | Asp | Arg | Glu | Phe |
|     |     |     |     | 635 |     |     |     |     |     | 640 |     |     |     | 645 |
| Asp | Leu | Leu | Arg | Ser | Asn | Arg | Leu | Glu | Glu | Arg | Ala | Ser | Ser | Asp |
|     |     |     |     | 650 |     |     |     |     |     | 655 |     |     |     | 660 |
| Lys | Asn | Val | Asp | Ser | Leu | Ser | Ala | Pro | His | Asn | His | Pro | Pro | Glu |
|     |     |     |     | 665 |     |     |     |     |     | 670 |     |     |     | 675 |
| Asp | Pro | Phe | Gly | Ser | Val | Pro | Phe | Ile | Ser | His | Ser | Gly | Ser | Pro |
|     |     |     |     | 680 |     |     |     |     |     | 685 |     |     |     | 690 |
| Glu | Lys | Lys | Ala | Glu | His | Ser | Ser | Ile | Asn | Gln | Glu | Asn | Gly | Thr |
|     |     |     |     | 695 |     |     |     |     |     | 700 |     |     |     | 705 |
| Ala | Asn | Pro | Ile | Lys | Asn | Gly | Lys | Thr | Ser | Pro | Ala | Ser | Lys | Asp |
|     |     |     |     | 710 |     |     |     |     |     | 715 |     |     |     | 720 |
| Gln | Arg | Thr | Gly | Lys | Lys | Thr | Ser | Val | Gln | Gly | Gln | Val | Gln | Lys |
|     |     |     |     | 725 |     |     |     |     |     | 730 |     |     |     | 735 |
| Gly | Asn | Asp | Glu | Ser | Glu | Ser | Asp | Phe | Glu | Ser | Asp | Pro | Pro | Ser |
|     |     |     |     | 740 |     |     |     |     |     | 745 |     |     |     | 750 |
| Pro | Lys | Ser | Ser | Glu | Glu | Glu | Gln | Asp | Asp | Glu | Glu | Val | Leu |     |
|     |     |     |     | 755 |     |     |     |     |     | 760 |     |     |     | 765 |
| Gln | Gly | Glu | Gln | Gly | Asp | Phe | Asn | Asp | Asp | Asp | Thr | Glu | Pro | Glu |
|     |     |     |     | 770 |     |     |     |     |     | 775 |     |     |     | 780 |
| Asn | Leu | Gly | His | Arg | Pro | Leu | Leu | Met | Asp | Ser | Glu | Asp | Glu | Glu |
|     |     |     |     | 785 |     |     |     |     |     | 790 |     |     |     | 795 |
| Glu | Glu | Glu | Lys | His | Ser | Ser | Asp | Ser | Asp | Tyr | Glu | Gln | Ala | Lys |
|     |     |     |     | 800 |     |     |     |     |     | 805 |     |     |     | 810 |
| Ala | Lys | Tyr | Ser | Asp | Met | Ser | Ser | Val | Tyr | Arg | Asp | Arg | Ser | Gly |
|     |     |     |     | 815 |     |     |     |     |     | 820 |     |     |     | 825 |
| Ser | Gly | Pro | Thr | Gln | Asp | Leu | Asn | Thr | Ile | Leu | Leu | Thr | Ser | Ala |
|     |     |     |     | 830 |     |     |     |     |     | 835 |     |     |     | 840 |

Gln Leu Ser Ser Asp Val Ala Val Glu Thr Pro Lys Gln Glu Phe  
 845 850 855  
 Asp Val Phe Gly Ala Val Pro Phe Phe Ala Val Arg Ala Gln Gln  
 860 865 870  
 Pro Gln Gln Glu Lys Asn Glu Lys Asn Leu Pro Gln His Arg Phe  
 875 880 885  
 Pro Ala Ala Gly Leu Glu Gln Glu Glu Phe Asp Val Phe Thr Lys  
 890 895 900  
 Ala Pro Phe Ser Lys Lys Val Asn Val Gln Glu Cys His Ala Val  
 905 910 915  
 Gly Pro Glu Ala His Thr Ile Pro Gly Tyr Pro Lys Ser Val Asp  
 920 925 930  
 Val Phe Gly Ser Thr Pro Phe Gln Pro Phe Leu Thr Ser Thr Ser  
 935 940 945  
 Lys Ser Glu Ser Asn Glu Asp Leu Phe Gly Leu Val Pro Phe Asp  
 950 955 960  
 Glu Ile Thr Gly Ser Gln Gln Gln Lys Val Lys Gln Arg Thr Leu  
 965 970 975  
 Gln Lys Leu Ser Ser Arg Gln Arg Arg Thr Lys Gln Asp Met Ser  
 980 985 990  
 Lys Ser Asn Gly Lys Arg His His Gly Thr Pro Thr Ser Thr Lys  
 995 1000 1005  
 Lys Thr Leu Lys Pro Thr Tyr Arg Thr Pro Glu Arg Ala Arg Arg  
 1010 1015 1020  
 His Lys Lys Val Gly Arg Arg Asp Ser Gln Ser Ser Asn Glu Phe  
 1025 1030 1035  
 Leu Thr Ile Ser Asp Ser Lys Glu Asn Ile Ser Val Ala Leu Thr  
 1040 1045 1050  
 Asp Gly Lys Asp Arg Gly Asn Val Leu Gln Pro Glu Glu Ser Leu  
 1055 1060 1065  
 Leu Asp Pro Phe Gly Ala Lys Pro Phe His Ser Pro Asp Leu Ser  
 1070 1075 1080  
 Trp His Pro Pro His Gln Gly Leu Ser Asp Ile Arg Ala Asp His  
 1085 1090 1095  
 Asn Thr Val Leu Pro Gly Arg Pro Arg Gln Asn Ser Leu His Gly  
 1100 1105 1110  
 Ser Phe His Ser Ala Asp Val Leu Lys Met Asp Asp Phe Gly Ala  
 1115 1120 1125  
 Val Pro Phe Thr Glu Leu Val Val Gln Ser Ile Thr Pro His Gln  
 1130 1135 1140  
 Ser Gln Gln Ser Gln Pro Val Glu Leu Asp Pro Phe Gly Ala Ala  
 1145 1150 1155  
 Pro Phe Pro Ser Lys Gln  
 1160

&lt;210&gt; 2

&lt;211&gt; 587

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3684162CD1

&lt;400&gt; 2

Met Ala Glu Lys Gly Asp Cys Ile Ala Ser Val Tyr Gly Tyr Asp

|   |     |     |     |
|---|-----|-----|-----|
| 1   | 5   | 10  | 15  |
| Leu Gly Gly Arg Phe Val Asp Phe Gln Pro Leu Gly Phe Gly Val |     |     |     |
|   | 20  | 25  | 30  |
| Asn Gly Leu Val Leu Ser Ala Val Asp Ser Arg Ala Cys Arg Lys |     |     |     |
|   | 35  | 40  | 45  |
| Val Ala Val Lys Lys Ile Ala Leu Ser Asp Ala Arg Ser Met Lys |     |     |     |
|   | 50  | 55  | 60  |
| His Ala Leu Arg Glu Ile Lys Ile Ile Arg Arg Leu Asp His Asp |     |     |     |
|   | 65  | 70  | 75  |
| Asn Ile Val Lys Val Tyr Glu Val Leu Gly Pro Lys Gly Thr Asp |     |     |     |
|   | 80  | 85  | 90  |
| Leu Gln Gly Glu Leu Phe Lys Phe Ser Val Ala Tyr Ile Val Gln |     |     |     |
|   | 95  | 100 | 105 |
| Glu Tyr Met Glu Thr Asp Leu Ala Arg Leu Leu Glu Gln Gly Thr |     |     |     |
|   | 110 | 115 | 120 |
| Leu Ala Glu Glu His Ala Lys Leu Phe Met Tyr Gln Leu Leu Arg |     |     |     |
|   | 125 | 130 | 135 |
| Gly Leu Lys Tyr Ile His Ser Ala Asn Val Leu His Arg Asp Leu |     |     |     |
|   | 140 | 145 | 150 |
| Lys Pro Ala Asn Ile Phe Ile Ser Thr Glu Asp Leu Val Leu Lys |     |     |     |
|   | 155 | 160 | 165 |
| Ile Gly Asp Phe Gly Leu Ala Arg Ile Val Asp Gln His Tyr Ser |     |     |     |
|   | 170 | 175 | 180 |
| His Lys Gly Tyr Leu Ser Glu Gly Leu Val Thr Lys Trp Tyr Arg |     |     |     |
|   | 185 | 190 | 195 |
| Ser Pro Arg Leu Leu Leu Ser Pro Asn Asn Tyr Thr Lys Ala Ile |     |     |     |
|   | 200 | 205 | 210 |
| Asp Met Trp Ala Ala Gly Cys Ile Leu Ala Glu Met Leu Thr Gly |     |     |     |
|   | 215 | 220 | 225 |
| Arg Met Leu Phe Ala Gly Ala His Glu Leu Glu Gln Met Gln Leu |     |     |     |
|   | 230 | 235 | 240 |
| Ile Leu Glu Thr Ile Pro Val Ile Arg Glu Glu Asp Lys Asp Glu |     |     |     |
|   | 245 | 250 | 255 |
| Leu Leu Arg Val Met Pro Ser Phe Val Ser Ser Thr Trp Glu Val |     |     |     |
|   | 260 | 265 | 270 |
| Lys Arg Pro Leu Arg Lys Leu Leu Pro Glu Val Asn Ser Glu Ala |     |     |     |
|   | 275 | 280 | 285 |
| Ile Asp Phe Leu Glu Lys Ile Leu Thr Phe Asn Pro Met Asp Arg |     |     |     |
|   | 290 | 295 | 300 |
| Leu Thr Ala Glu Met Gly Leu Gln His Pro Tyr Met Ser Pro Tyr |     |     |     |
|   | 305 | 310 | 315 |
| Ser Cys Pro Glu Asp Glu Pro Thr Ser Gln His Pro Phe Arg Ile |     |     |     |
|   | 320 | 325 | 330 |
| Glu Asp Glu Ile Asp Asp Ile Val Leu Met Ala Ala Asn Gln Ser |     |     |     |
|   | 335 | 340 | 345 |
| Gln Leu Ser Asn Trp Asp Thr Cys Ser Ser Arg Tyr Pro Val Ser |     |     |     |
|   | 350 | 355 | 360 |
| Leu Ser Ser Asp Leu Glu Trp Arg Pro Asp Arg Cys Gln Asp Ala |     |     |     |
|   | 365 | 370 | 375 |
| Ser Glu Val Gln Arg Asp Pro Arg Ala Gly Ser Ala Pro Leu Ala |     |     |     |
|   | 380 | 385 | 390 |
| Glu Asp Val Gln Val Asp Pro Arg Lys Asp Ser His Ser Ser Ser |     |     |     |
|   | 395 | 400 | 405 |
| Glu Arg Phe Leu Glu Gln Ser His Ser Ser Met Glu Arg Ala Phe |     |     |     |
|   | 410 | 415 | 420 |
| Glu Ala Asp Tyr Gly Arg Ser Cys Asp Tyr Lys Val Gly Ser Pro |     |     |     |



|                 |                     |                         |     |  |     |
|-----------------|---------------------|-------------------------|-----|--|-----|
|                 | 425                 |                         | 430 |  | 435 |
| Ser Tyr Leu Asp | Lys Leu Leu Trp Arg | Asp Asn Lys Pro His His |     |  |     |
|                 | 440                 |                         | 445 |  | 450 |
| Tyr Ser Glu Pro | Lys Leu Ile Leu Asp | Leu Ser His Trp Lys Gln |     |  |     |
|                 | 455                 |                         | 460 |  | 465 |
| Ala Ala Gly Ala | Pro Pro Thr Ala Thr | Gly Leu Ala Asp Thr Gly |     |  |     |
|                 | 470                 |                         | 475 |  | 480 |
| Ala Arg Glu Asp | Glu Pro Ala Ser Leu | Phe Leu Glu Ile Ala Gln |     |  |     |
|                 | 485                 |                         | 490 |  | 495 |
| Trp Val Lys Ser | Thr Gln Gly Gly Pro | Glu His Ala Ser Pro Pro |     |  |     |
|                 | 500                 |                         | 505 |  | 510 |
| Ala Asp Asp Pro | Glu Arg Arg Leu Ser | Ala Ser Pro Pro Gly Arg |     |  |     |
|                 | 515                 |                         | 520 |  | 525 |
| Pro Ala Pro Val | Asp Gly Gly Ala Ser | Pro Gln Phe Asp Leu Asp |     |  |     |
|                 | 530                 |                         | 535 |  | 540 |
| Val Phe Ile Ser | Arg Ala Leu Lys Leu | Cys Thr Lys Pro Glu Asp |     |  |     |
|                 | 545                 |                         | 550 |  | 555 |
| Leu Pro Asp Asn | Lys Leu Gly Asp Leu | Asn Gly Ala Cys Ile Pro |     |  |     |
|                 | 560                 |                         | 565 |  | 570 |
| Glu His Pro Gly | Asp Leu Val Gln Thr | Glu Ala Phe Ser Lys Glu |     |  |     |
|                 | 575                 |                         | 580 |  | 585 |
| Arg Trp         |                     |                         |     |  |     |

&lt;210&gt; 3

&lt;211&gt; 1275

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3736769CD1

&lt;400&gt; 3.

|                 |                     |                         |
|-----------------|---------------------|-------------------------|
| Met Glu Asn Phe | Ile Leu Tyr Glu Glu | Ile Gly Arg Gly Ser Lys |
| 1               | 5                   | 10 15                   |
| Thr Val Val Tyr | Lys Gly Arg Arg Lys | Gly Thr Ile Asn Phe Val |
|                 | 20                  | 25 30                   |
| Ala Ile Leu Cys | Thr Asp Lys Cys Arg | Arg Pro Glu Ile Thr Asn |
|                 | 35                  | 40 45                   |
| Trp Val Arg Leu | Thr Arg Glu Ile Lys | His Lys Asn Ile Val Thr |
|                 | 50                  | 55 60                   |
| Phe His Glu Trp | Tyr Glu Thr Ser Asn | His Leu Trp Leu Val Val |
|                 | 65                  | 70 75                   |
| Glu Leu Cys Thr | Gly Gly Ser Leu Lys | Thr Val Ile Ala Gln Asp |
|                 | 80                  | 85 90                   |
| Glu Asn Leu Pro | Glu Asp Val Val Arg | Glu Phe Gly Ile Asp Leu |
|                 | 95                  | 100 105                 |
| Ile Ser Gly Leu | His His Leu His Lys | Leu Gly Ile Leu Phe Cys |
|                 | 110                 | 115 120                 |
| Asp Ile Ser Pro | Arg Lys Ile Leu Leu | Glu Gly Pro Gly Thr Leu |
|                 | 125                 | 130 135                 |
| Lys Phe Ser Asn | Phe Cys Leu Ala Lys | Val Glu Gly Glu Asn Leu |
|                 | 140                 | 145 150                 |
| Glu Glu Phe Phe | Ala Leu Val Ala Ala | Glu Glu Gly Gly Gly Asp |
|                 | 155                 | 160 165                 |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Asn | Gly | Glu | Asn | Val | Leu | Lys | Lys | Ser | Met | Lys | Ser | Arg | Val | Lys |
|     |     |     |     |     | 170 |     |     |     | 175 |     |     |     |     | 180 |
| Gly | Ser | Pro | Val | Tyr | Thr | Ala | Pro | Glu | Val | Val | Arg | Gly | Ala | Asp |
|     |     |     |     |     | 185 |     |     |     | 190 |     |     |     |     | 195 |
| Phe | Ser | Ile | Ser | Ser | Asp | Leu | Trp | Ser | Leu | Gly | Cys | Leu | Leu | Tyr |
|     |     |     |     |     | 200 |     |     |     | 205 |     |     |     |     | 210 |
| Glu | Met | Phe | Ser | Gly | Lys | Pro | Pro | Phe | Phe | Ser | Glu | Ser | Val | Ser |
|     |     |     |     |     | 215 |     |     |     | 220 |     |     |     |     | 225 |
| Glu | Leu | Thr | Glu | Lys | Ile | Leu | Cys | Glu | Asp | Pro | Leu | Pro | Pro | Ile |
|     |     |     |     |     | 230 |     |     |     | 235 |     |     |     |     | 240 |
| Pro | Lys | Asp | Ser | Ser | Arg | Pro | Lys | Ala | Ser | Ser | Asp | Phe | Ile | Asn |
|     |     |     |     |     | 245 |     |     |     | 250 |     |     |     |     | 255 |
| Leu | Leu | Asp | Gly | Leu | Leu | Gln | Arg | Asp | Pro | Gln | Lys | Arg | Leu | Thr |
|     |     |     |     |     | 260 |     |     |     | 265 |     |     |     |     | 270 |
| Trp | Thr | Arg | Leu | Leu | Gln | His | Ser | Phe | Trp | Lys | Lys | Ala | Phe | Ala |
|     |     |     |     |     | 275 |     |     |     | 280 |     |     |     |     | 285 |
| Gly | Ala | Asp | Gln | Glu | Ser | Ser | Val | Glu | Asp | Leu | Ser | Leu | Ser | Arg |
|     |     |     |     |     | 290 |     |     |     | 295 |     |     |     |     | 300 |
| Asn | Thr | Met | Glu | Cys | Ser | Gly | Pro | Gln | Asp | Ser | Lys | Glu | Leu | Leu |
|     |     |     |     |     | 305 |     |     |     | 310 |     |     |     |     | 315 |
| Gln | Asn | Ser | Gln | Ser | Arg | Gln | Ala | Lys | Gly | His | Lys | Ser | Gly | Gln |
|     |     |     |     |     | 320 |     |     |     | 325 |     |     |     |     | 330 |
| Pro | Leu | Gly | His | Ser | Phe | Arg | Leu | Glu | Asn | Pro | Thr | Glu | Phe | Arg |
|     |     |     |     |     | 335 |     |     |     | 340 |     |     |     |     | 345 |
| Pro | Lys | Ser | Thr | Leu | Glu | Gly | Gln | Leu | Asn | Glu | Ser | Met | Phe | Leu |
|     |     |     |     |     | 350 |     |     |     | 355 |     |     |     |     | 360 |
| Leu | Ser | Ser | Arg | Pro | Thr | Pro | Arg | Thr | Ser | Thr | Ala | Val | Glu | Val |
|     |     |     |     |     | 365 |     |     |     | 370 |     |     |     |     | 375 |
| Ser | Pro | Gly | Glu | Asp | Met | Thr | His | Cys | Ser | Pro | Gln | Lys | Thr | Ser |
|     |     |     |     |     | 380 |     |     |     | 385 |     |     |     |     | 390 |
| Pro | Leu | Thr | Lys | Ile | Thr | Ser | Gly | His | Leu | Ser | Gln | Gln | Asp | Leu |
|     |     |     |     |     | 395 |     |     |     | 400 |     |     |     |     | 405 |
| Glu | Ser | Gln | Met | Arg | Glu | Leu | Ile | Tyr | Thr | Asp | Ser | Asp | Leu | Val |
|     |     |     |     |     | 410 |     |     |     | 415 |     |     |     |     | 420 |
| Val | Thr | Pro | Ile | Ile | Asp | Asn | Pro | Lys | Ile | Met | Lys | Gln | Pro | Pro |
|     |     |     |     |     | 425 |     |     |     | 430 |     |     |     |     | 435 |
| Val | Lys | Phe | Asp | Ala | Lys | Ile | Leu | His | Leu | Pro | Thr | Tyr | Ser | Val |
|     |     |     |     |     | 440 |     |     |     | 445 |     |     |     |     | 450 |
| Asp | Lys | Leu | Leu | Phe | Leu | Lys | Asp | Gln | Asp | Trp | Asn | Asp | Phe | Leu |
|     |     |     |     |     | 455 |     |     |     | 460 |     |     |     |     | 465 |
| Gln | Gln | Val | Cys | Ser | Gln | Ile | Asp | Ser | Thr | Glu | Lys | Ser | Met | Gly |
|     |     |     |     |     | 470 |     |     |     | 475 |     |     |     |     | 480 |
| Ala | Ser | Arg | Ala | Lys | Leu | Asn | Leu | Leu | Cys | Tyr | Leu | Cys | Val | Val |
|     |     |     |     |     | 485 |     |     |     | 490 |     |     |     |     | 495 |
| Ala | Gly | His | Gln | Glu | Val | Ala | Thr | Arg | Leu | Leu | His | Ser | Pro | Leu |
|     |     |     |     |     | 500 |     |     |     | 505 |     |     |     |     | 510 |
| Phe | Gln | Leu | Leu | Ile | Gln | His | Leu | Arg | Ile | Ala | Pro | Asn | Trp | Asp |
|     |     |     |     |     | 515 |     |     |     | 520 |     |     |     |     | 525 |
| Ile | Arg | Ala | Lys | Val | Ala | His | Val | Ile | Gly | Leu | Leu | Ala | Ser | His |
|     |     |     |     |     | 530 |     |     |     | 535 |     |     |     |     | 540 |
| Thr | Thr | Glu | Leu | Gln | Glu | Asn | Thr | Pro | Val | Val | Glu | Ala | Ile | Val |
|     |     |     |     |     | 545 |     |     |     | 550 |     |     |     |     | 555 |
| Leu | Leu | Thr | Glu | Leu | Ile | Arg | Glu | Asn | Phe | Arg | Asn | Ser | Lys | Leu |
|     |     |     |     |     | 560 |     |     |     | 565 |     |     |     |     | 570 |
| Lys | Gln | Cys | Leu | Leu | Pro | Thr | Leu | Gly | Glu | Leu | Ile | Tyr | Leu | Val |
|     |     |     |     |     | 575 |     |     |     | 580 |     |     |     |     | 585 |

|   |     |      |      |
|---|-----|------|------|
| Ala Thr Gln Glu Glu Lys Lys Lys Asn Pro Arg Glu Cys Trp Ala | 590 | 595  | 600  |
| Val Pro Leu Ala Ala Tyr Thr Val Leu Met Arg Cys Leu Arg Glu | 605 | 610  | 615  |
| Gly Glu Glu Arg Val Val Asn His Met Ala Ala Lys Ile Ile Glu | 620 | 625  | 630  |
| Asn Val Cys Thr Thr Phe Ser Ala Gln Ser Gln Gly Phe Ile Thr | 635 | 640  | 645  |
| Gly Glu Ile Gly Pro Ile Leu Trp Tyr Leu Phe Arg His Ser Thr | 650 | 655  | 660  |
| Ala Asp Ser Leu Arg Ile Thr Ala Val Ser Ala Leu Cys Arg Ile | 665 | 670  | 675  |
| Thr Arg His Ser Pro Thr Ala Phe Gln Asn Val Ile Glu Lys Val | 680 | 685  | 690  |
| Gly Leu Asn Ser Val Ile Asn Ser Leu Ala Ser Ala Ile Cys Lys | 695 | 700  | 705  |
| Val Gln Gln Tyr Met Leu Thr Leu Phe Ala Ala Met Leu Ser Cys | 710 | 715  | 720  |
| Gly Ile His Leu Gln Arg Leu Ile Gln Glu Lys Gly Phe Val Ser | 725 | 730  | 735  |
| Thr Ile Ile Arg Leu Leu Asp Ser Pro Ser Thr Cys Ile Arg Ala | 740 | 745  | 750  |
| Lys Ala Phe Leu Val Leu Leu Tyr Ile Leu Ile Tyr Asn Arg Glu | 755 | 760  | 765  |
| Met Leu Leu Leu Ser Cys Gln Ala Arg Leu Val Met Tyr Ile Glu | 770 | 775  | 780  |
| Arg Asp Ser Arg Lys Thr Thr Pro Gly Lys Glu Gln Gln Ser Gly | 785 | 790  | 795  |
| Asn Glu Tyr Leu Ser Lys Cys Leu Asp Leu Leu Ile Cys His Ile | 800 | 805  | 810  |
| Val Gln Glu Leu Pro Arg Ile Leu Gly Asp Ile Leu Asn Ser Leu | 815 | 820  | 825  |
| Ala Asn Val Ser Gly Arg Lys His Pro Ser Thr Val Gln Val Lys | 830 | 835  | 840  |
| Gln Leu Lys Leu Cys Leu Pro Leu Met Pro Val Val Leu His Leu | 845 | 850  | 855  |
| Val Thr Ser Gln Val Phe Arg Pro Gln Val Val Thr Glu Glu Phe | 860 | 865  | 870  |
| Leu Phe Ser Tyr Gly Thr Ile Leu Ser His Ile Lys Ser Val Asp | 875 | 880  | 885  |
| Ser Gly Glu Thr Asn Ile Asp Gly Ala Ile Gly Leu Thr Ala Ser | 890 | 895  | 900  |
| Glu Glu Phe Ile Lys Ile Thr Leu Ser Ala Phe Glu Ala Ile Ile | 905 | 910  | 915  |
| Gln Tyr Pro Ile Leu Leu Lys Asp Tyr Arg Ser Thr Val Val Asp | 920 | 925  | 930  |
| Tyr Ile Leu Pro Pro Leu Val Ser Leu Val Gln Ser Gln Asn Val | 935 | 940  | 945  |
| Glu Trp Arg Leu Phe Ser Leu Arg Leu Leu Ser Glu Thr Thr Ser | 950 | 955  | 960  |
| Leu Leu Val Asn Gln Glu Phe Gly Asp Gly Lys Glu Lys Ala Ser | 965 | 970  | 975  |
| Val Asp Ser Asp Ser Asn Leu Leu Ala Leu Ile Arg Asp Val Leu | 980 | 985  | 990  |
| Leu Pro Gln Tyr Glu His Ile Leu Leu Glu Pro Asp Pro Val Pro | 995 | 1000 | 1005 |

Ala Tyr Ala Leu Lys Leu Leu Val Ala Met Thr Glu His Asn Pro  
 1010 1015 1020  
 Thr Phe Thr Arg Leu Val Glu Glu Ser Lys Leu Ile Pro Leu Ile  
 1025 1030 1035  
 Phe Glu Val Thr Leu Glu His Gln Glu Ser Ile Leu Gly Asn Thr  
 1040 1045 1050  
 Met Gln Ser Val Ile Ala Leu Leu Ser Asn Leu Val Ala Cys Lys  
 1055 1060 1065  
 Asp Ser Asn Met Glu Leu Leu Tyr Glu Gln Gly Leu Val Ser His  
 1070 1075 1080  
 Ile Cys Asn Leu Leu Thr Glu Thr Ala Thr Leu Cys Leu Asp Val  
 1085 1090 1095  
 Asp Asn Lys Asn Asn Asn Glu Met Ala Ala Pro Leu Leu Phe Ser  
 1100 1105 1110  
 Leu Leu Asp Ile Leu His Ser Met Leu Thr Tyr Thr Ser Gly Ile  
 1115 1120 1125  
 Val Arg Leu Ala Leu Gln Ala Gln Lys Ser Gly Ser Gly Glu Asp  
 1130 1135 1140  
 Pro Gln Ala Ala Glu Asp Leu Leu Leu Leu Asn Arg Pro Leu Thr  
 1145 1150 1155  
 Asp Leu Ile Ser Leu Leu Ile Pro Leu Leu Pro Asn Glu Asp Pro  
 1160 1165 1170  
 Glu Ile Phe Asp Val Ser Ser Lys Cys Leu Ser Ile Leu Val Gln  
 1175 1180 1185  
 Leu Tyr Gly Gly Glu Asn Pro Asp Ser Leu Ser Pro Glu Asn Val  
 1190 1195 1200  
 Glu Ile Phe Ala His Leu Leu Thr Ser Lys Glu Asp Pro Lys Glu  
 1205 1210 1215  
 Gln Lys Leu Leu Leu Arg Ile Leu Arg Arg Met Ile Thr Ser Asn  
 1220 1225 1230  
 Glu Lys His Leu Glu Ser Leu Lys Asn Ala Gly Ser Leu Leu Arg  
 1235 1240 1245  
 Ala Leu Glu Arg Leu Ala Pro Gly Ser Gly Ser Phe Ala Asp Ser  
 1250 1255 1260  
 Ala Val Ala Pro Leu Ala Leu Glu Ile Leu Gln Ala Val Gly His  
 1265 1270 1275

&lt;210&gt; 4

&lt;211&gt; 1406

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7474632CD1

&lt;400&gt; 4

Met His Gln Thr Leu Cys Leu Asn Pro Glu Ser Leu Lys Met Ser  
 1 5 10 15  
 Ala Cys Ser Asp Phe Val Glu His Ile Trp Lys Pro Gly Ser Cys  
 20 25 30  
 Lys Asn Cys Phe Cys Leu Arg Ser Asp His Gln Leu Val Ala Gly  
 35 40 45  
 Pro Pro Gln Pro Arg Ala Gly Ser Leu Pro Pro Pro Arg Leu  
 50 55 60

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Pro | Pro | Arg | Pro | Glu | Asn | Cys | Arg | Leu | Glu | Asp | Glu | Gly | Val | Asn | 65  | 70  | 75  |
| Ser | Ser | Pro | Tyr | Ser | Lys | Pro | Thr | Ile | Ala | Val | Lys | Pro | Thr | Met | 80  | 85  | 90  |
| Met | Ser | Ser | Glu | Ala | Ser | Asp | Val | Trp | Thr | Glu | Ala | Asn | Leu | Ser | 95  | 100 | 105 |
| Ala | Glu | Val | Ser | Gln | Val | Ile | Trp | Arg | Arg | Ala | Pro | Gly | Lys | Leu | 110 | 115 | 120 |
| Pro | Leu | Pro | Lys | Gln | Glu | Asp | Ala | Pro | Val | Val | Tyr | Leu | Gly | Ser | 125 | 130 | 135 |
| Phe | Arg | Gly | Val | Gln | Lys | Pro | Ala | Gly | Pro | Ser | Thr | Ser | Pro | Asp | 140 | 145 | 150 |
| Gly | Asn | Ser | Arg | Cys | Pro | Pro | Ala | Tyr | Thr | Met | Val | Gly | Leu | His | 155 | 160 | 165 |
| Asn | Leu | Glu | Pro | Arg | Gly | Glu | Arg | Asn | Ile | Ala | Phe | His | Pro | Val | 170 | 175 | 180 |
| Ser | Phe | Pro | Glu | Glu | Lys | Ala | Val | His | Lys | Glu | Lys | Pro | Ser | Phe | 185 | 190 | 195 |
| Pro | Tyr | Gln | Asp | Arg | Pro | Ser | Thr | Gln | Glu | Ser | Phe | Arg | Gln | Lys | 200 | 205 | 210 |
| Leu | Ala | Ala | Phe | Ala | Gly | Thr | Thr | Ser | Gly | Cys | His | Gln | Gly | Pro | 215 | 220 | 225 |
| Gly | Pro | Leu | Arg | Glu | Ser | Leu | Pro | Ser | Glu | Asp | Asp | Ser | Asp | Gln | 230 | 235 | 240 |
| Arg | Cys | Ser | Pro | Ser | Gly | Asp | Ser | Glu | Gly | Gly | Glu | Tyr | Cys | Ser | 245 | 250 | 255 |
| Ile | Leu | Asp | Cys | Cys | Pro | Gly | Ser | Pro | Val | Ala | Lys | Ala | Ala | Ser | 260 | 265 | 270 |
| Gln | Thr | Ala | Gly | Ser | Arg | Gly | Arg | His | Gly | Gly | Arg | Asp | Cys | Ser | 275 | 280 | 285 |
| Pro | Thr | Cys | Trp | Glu | Gln | Gly | Lys | Cys | Ser | Gly | Pro | Ala | Glu | Gln | 290 | 295 | 300 |
| Glu | Lys | Arg | Gly | Pro | Ser | Phe | Pro | Lys | Glu | Cys | Cys | Ser | Gln | Gly | 305 | 310 | 315 |
| Pro | Thr | Ala | His | Pro | Ser | Cys | Leu | Gly | Pro | Lys | Lys | Leu | Ser | Leu | 320 | 325 | 330 |
| Thr | Ser | Glu | Ala | Ala | Ile | Ser | Ser | Asp | Gly | Leu | Ser | Cys | Gly | Ser | 335 | 340 | 345 |
| Gly | Ser | Gly | Ser | Gly | Ser | Gly | Ala | Ser | Ser | Pro | Phe | Val | Pro | His | 350 | 355 | 360 |
| Leu | Glu | Ser | Asp | Tyr | Cys | Ser | Leu | Met | Lys | Glu | Pro | Ala | Pro | Glu | 365 | 370 | 375 |
| Lys | Gln | Gln | Asp | Pro | Gly | Cys | Pro | Gly | Val | Thr | Pro | Ser | Arg | Cys | 380 | 385 | 390 |
| Leu | Gly | Leu | Thr | Gly | Glu | Pro | Gln | Pro | Pro | Ala | Gln | Pro | Gln | Glu | 395 | 400 | 405 |
| Ala | Thr | Gln | Pro | Glu | Pro | Ile | Tyr | Ala | Glu | Ser | Thr | Lys | Arg | Lys | 410 | 415 | 420 |
| Lys | Ala | Ala | Pro | Val | Pro | Ser | Lys | Ser | Gln | Ala | Lys | Ile | Glu | His | 425 | 430 | 435 |
| Ala | Ala | Ala | Ala | Gln | Gly | Gln | Gly | Gln | Val | Cys | Thr | Gly | Asn | Ala | 440 | 445 | 450 |
| Trp | Ala | Gln | Lys | Ala | Ala | Ser | Gly | Trp | Gly | Arg | Asp | Ser | Pro | Asp | 455 | 460 | 465 |
| Pro | Thr | Pro | Gln | Val | Ser | Ala | Thr | Ile | Thr | Val | Met | Ala | Ala | His | 470 | 475 | 480 |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Pro | Glu | Glu | Asp | His | Arg | Thr | Ile | Tyr | Leu | Ser | Ser | Pro | Asp | Ser | 485 | 490 | 495 |
| Ala | Val | Gly | Val | Gln | Trp | Pro | Arg | Gly | Pro | Val | Ser | Gln | Asn | Ser | 500 | 505 | 510 |
| Glu | Val | Gly | Glu | Glu | Glu | Thr | Ser | Ala | Gly | Gln | Gly | Leu | Ser | Ser | 515 | 520 | 525 |
| Arg | Glu | Ser | His | Ala | His | Ser | Ala | Ser | Glu | Ser | Lys | Pro | Lys | Glu | 530 | 535 | 540 |
| Arg | Pro | Ala | Ile | Pro | Pro | Lys | Leu | Ser | Lys | Ser | Ser | Pro | Val | Gly | 545 | 550 | 555 |
| Ser | Pro | Val | Ser | Pro | Ser | Ala | Gly | Gly | Pro | Pro | Val | Ser | Pro | Leu | 560 | 565 | 570 |
| Ala | Asp | Leu | Ser | Asp | Gly | Ser | Ser | Gly | Gly | Ser | Ser | Ile | Gly | Pro | 575 | 580 | 585 |
| Gln | Pro | Pro | Ser | Gln | Gly | Pro | Ala | Asp | Pro | Ala | Pro | Ser | Cys | Arg | 590 | 595 | 600 |
| Thr | Asn | Gly | Val | Ala | Ile | Ser | Asp | Pro | Ser | Arg | Cys | Pro | Gln | Pro | 605 | 610 | 615 |
| Ala | Ala | Ser | Ser | Ala | Ser | Glu | Gln | Arg | Arg | Pro | Arg | Phe | Gln | Ala | 620 | 625 | 630 |
| Gly | Thr | Trp | Ser | Arg | Gln | Cys | Arg | Ile | Glu | Glu | Glu | Glu | Glu | Val | 635 | 640 | 645 |
| Glu | Gln | Glu | Leu | Leu | Ser | His | Ser | Trp | Gly | Arg | Glu | Thr | Lys | Asn | 650 | 655 | 660 |
| Gly | Pro | Thr | Asp | His | Ser | Asn | Ser | Thr | Thr | Trp | His | Arg | Leu | His | 665 | 670 | 675 |
| Pro | Thr | Asp | Gly | Ser | Ser | Gly | Gln | Asn | Ser | Lys | Val | Gly | Thr | Gly | 680 | 685 | 690 |
| Met | Ser | Lys | Ser | Ala | Ser | Phe | Ala | Phe | Glu | Phe | Pro | Lys | Asp | Arg | 695 | 700 | 705 |
| Ser | Gly | Ile | Glu | Thr | Phe | Ser | Pro | Pro | Pro | Pro | Pro | Lys | Ser |     | 710 | 715 | 720 |
| Arg | His | Leu | Leu | Lys | Met | Asn | Lys | Ser | Ser | Ser | Asp | Leu | Glu | Lys | 725 | 730 | 735 |
| Val | Ser | Gln | Gly | Ser | Ala | Glu | Ser | Leu | Ser | Pro | Ser | Phe | Arg | Gly | 740 | 745 | 750 |
| Val | His | Val | Ser | Phe | Thr | Thr | Gly | Ser | Thr | Asp | Ser | Leu | Ala | Ser | 755 | 760 | 765 |
| Asp | Ser | Arg | Thr | Cys | Ser | Asp | Gly | Gly | Pro | Ser | Ser | Glu | Leu | Ala | 770 | 775 | 780 |
| His | Ser | Pro | Thr | Asn | Ser | Gly | Lys | Lys | Leu | Phe | Ala | Pro | Val | Pro | 785 | 790 | 795 |
| Phe | Pro | Ser | Gly | Ser | Thr | Glu | Asp | Val | Ser | Pro | Ser | Gly | Pro | Gln | 800 | 805 | 810 |
| Gln | Pro | Pro | Pro | Leu | Pro | Gln | Lys | Lys | Ile | Val | Ser | Arg | Ala | Ala | 815 | 820 | 825 |
| Ser | Ser | Pro | Asp | Gly | Phe | Phe | Trp | Thr | Gln | Gly | Ser | Pro | Lys | Pro | 830 | 835 | 840 |
| Gly | Thr | Ala | Ser | Pro | Lys | Leu | Asn | Leu | Ser | His | Ser | Glu | Thr | Asn | 845 | 850 | 855 |
| Val | His | Asp | Glu | Ser | His | Phe | Ser | Tyr | Ser | Leu | Ser | Pro | Gly | Asn | 860 | 865 | 870 |
| Arg | His | His | Pro | Val | Phe | Ser | Ser | Ser | Asp | Pro | Leu | Glu | Lys | Ala | 875 | 880 | 885 |
| Phe | Lys | Gly | Ser | Gly | His | Trp | Leu | Pro | Ala | Ala | Gly | Leu | Ala | Gly | 890 | 895 | 900 |

|   |      |      |      |
|---|------|------|------|
| Asn Arg Gly Gly Cys Gly Ser Pro Gly Leu Gln Cys Lys Gly Ala | 905  | 910  | 915  |
| Pro Ser Ala Ser Ser Ser Gln Leu Ser Val Ser Ser Gln Ala Ser | 920  | 925  | 930  |
| Thr Gly Ser Thr Gln Leu Gln Leu His Gly Leu Leu Ser Asn Ile | 935  | 940  | 945  |
| Ser Ser Lys Glu Gly Thr Tyr Ala Lys Leu Gly Gly Leu Tyr Thr | 950  | 955  | 960  |
| Gln Ser Leu Ala Arg Leu Val Ala Lys Cys Glu Asp Leu Phe Met | 965  | 970  | 975  |
| Gly Gly Gln Lys Lys Glu Leu His Phe Asn Glu Asn Asn Trp Ser | 980  | 985  | 990  |
| Leu Phe Lys Leu Thr Cys Asn Lys Pro Cys Cys Asp Ser Gly Asp | 995  | 1000 | 1005 |
| Ala Ile Tyr Tyr Cys Ala Thr Cys Ser Glu Asp Pro Gly Ser Thr | 1010 | 1015 | 1020 |
| Tyr Ala Val Lys Ile Cys Lys Ala Pro Glu Pro Lys Thr Val Ser | 1025 | 1030 | 1035 |
| Tyr Cys Ser Pro Ser Val Pro Val His Phe Asn Ile Gln Gln Asp | 1040 | 1045 | 1050 |
| Cys Gly His Phe Val Ala Ser Val Pro Ser Ser Met Leu Ser Ser | 1055 | 1060 | 1065 |
| Pro Asp Ala Pro Lys Asp Pro Val Pro Ala Leu Pro Thr His Pro | 1070 | 1075 | 1080 |
| Pro Ala Gln Glu Gln Asp Cys Val Val Val Ile Thr Arg Glu Val | 1085 | 1090 | 1095 |
| Pro His Gln Thr Ala Ser Asp Phe Val Arg Asp Ser Ala Ala Ser | 1100 | 1105 | 1110 |
| His Gln Ala Glu Pro Glu Ala Tyr Glu Arg Arg Val Cys Phe Leu | 1115 | 1120 | 1125 |
| Leu Leu Gln Leu Cys Asn Gly Leu Glu His Leu Lys Glu His Gly | 1130 | 1135 | 1140 |
| Ile Ile His Arg Asp Leu Cys Leu Glu Asn Leu Leu Leu Val His | 1145 | 1150 | 1155 |
| Cys Thr Leu Gln Ala Gly Pro Gly Pro Ala Pro Ala Pro Ala Pro | 1160 | 1165 | 1170 |
| Ala Pro Ala Pro Ala Ala Ala Ala Pro Pro Cys Ser Ser Ala Ala | 1175 | 1180 | 1185 |
| Pro Pro Ala Gly Gly Thr Leu Ser Pro Ala Ala Gly Pro Ala Ser | 1190 | 1195 | 1200 |
| Pro Glu Gly Pro Arg Glu Lys Gln Leu Pro Arg Leu Ile Ile Ser | 1205 | 1210 | 1215 |
| Asn Phe Leu Lys Ala Lys Gln Lys Pro Gly Gly Thr Pro Asn Leu | 1220 | 1225 | 1230 |
| Gln Gln Lys Lys Ser Gln Ala Arg Leu Ala Pro Glu Ile Val Ser | 1235 | 1240 | 1245 |
| Ala Ser Gln Tyr Arg Lys Phe Asp Glu Phe Gln Thr Gly Ile Leu | 1250 | 1255 | 1260 |
| Ile Tyr Gln Leu Leu His Gln Pro Asn Pro Phe Glu Val Arg Ala | 1265 | 1270 | 1275 |
| Gln Leu Arg Glu Arg Asp Tyr Arg Gln Glu Asp Leu Pro Pro Leu | 1280 | 1285 | 1290 |
| Pro Ala Leu Ser Leu Tyr Ser Pro Gly Leu Gln Gln Leu Ala His | 1295 | 1300 | 1305 |
| Leu Leu Leu Glu Ala Asp Pro Ile Lys Arg Ile Arg Ile Gly Glu | 1310 | 1315 | 1320 |

Ala Lys Arg Val Leu Gln Cys Leu Leu Trp Gly Pro Arg Arg Glu  
 1325 1330 1335  
 Leu Val Gln Gln Pro Gly Thr Ser Glu Glu Ala Leu Cys Gly Thr  
 1340 1345 1350  
 Leu His Asn Trp Ile Asp Met Lys Arg Ala Leu Met Met Met Lys  
 1355 1360 1365  
 Phe Ala Glu Lys Ala Val Asp Arg Arg Arg Gly Val Glu Leu Glu  
 1370 1375 1380  
 Asp Trp Leu Cys Cys Gln Tyr Leu Ala Ser Ala Glu Pro Gly Ala  
 1385 1390 1395  
 Leu Leu Gln Ser Leu Lys Leu Leu Gln Leu Leu  
 1400 1405

&lt;210&gt; 5

&lt;211&gt; 463

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7472696CD1

&lt;400&gt; 5

Met Leu His Arg Glu Gly Pro Pro Pro Arg Pro Arg Leu Leu Ala  
 1 5 10 15  
 Arg Val Thr Leu Pro Pro Ala Ile Arg Ala Ala Ala Leu Gly Ala  
 20 25 30  
 Ser Phe Leu Thr Ser His Pro Ala Arg Ser Pro Glu Arg Ala Ser  
 35 40 45  
 Ala Ala Cys Arg Val Arg Pro Gly Leu Gly Ala Val Ala Arg Gly  
 50 55 60  
 Arg Ala Arg Gly Glu Ala Arg Leu Pro Arg Ser Ala Ser Ser Pro  
 65 70 75  
 Ala Pro Pro Thr Pro Gln Ala Gln Ala Pro Gln Thr Arg Ser Ser  
 80 85 90  
 Leu Arg Ser Pro Ser Pro Pro Ala Ser Arg Pro His Pro Phe Arg  
 95 100 105  
 Ala Pro Arg Arg Arg Gln Thr Thr Ser Asp Pro Pro Pro Pro Pro  
 110 115 120  
 Gly Tyr Arg Pro Gly Gln Pro Ala Arg Glu Gly Gly Arg Glu Leu  
 125 130 135  
 Pro Phe Cys Phe Ser His Leu Leu Val Asp Ile Pro Ala Pro Pro  
 140 145 150  
 Ala Pro Phe Asp His Arg Ile Val Thr Ala Lys Gln Gly Ala Val  
 155 160 165  
 Asn Ser Phe Tyr Thr Val Ser Lys Thr Glu Ile Leu Gly Gly Gly  
 170 175 180  
 Arg Phe Gly Gln Val His Lys Cys Glu Glu Thr Ala Thr Gly Leu  
 185 190 195  
 Lys Leu Ala Ala Lys Ile Ile Lys Thr Arg Gly Met Lys Asp Lys  
 200 205 210  
 Glu Glu Val Lys Asn Glu Ile Ser Val Met Asn Gln Leu Asp His  
 215 220 225  
 Ala Asn Leu Ile Gln Leu Tyr Asp Ala Phe Glu Ser Lys Asn Asp  
 230 235 240  
 Ile Val Leu Val Met Glu Tyr Val Asp Gly Gly Glu Leu Phe Asp



|   |     |     |
|---|-----|-----|
| 245   | 250 | 255 |
| Arg Ile Ile Asp Glu Ser Tyr Asn Leu Thr Glu Leu Asp Thr Ile |     |     |
| 260   | 265 | 270 |
| Leu Phe Met Lys Gln Ile Cys Glu Gly Ile Arg His Met His Gln |     |     |
| 275   | 280 | 285 |
| Met Tyr Ile Leu His Leu Asp Leu Lys Pro Glu Asn Ile Leu Cys |     |     |
| 290   | 295 | 300 |
| Val Asn Arg Asp Ala Glu Gln Ile Lys Ile Ile Asp Phe Gly Leu |     |     |
| 305   | 310 | 315 |
| Ala Arg Arg Tyr Lys Pro Arg Glu Lys Leu Lys Val Asn Phe Gly |     |     |
| 320   | 325 | 330 |
| Thr Pro Glu Phe Leu Ala Pro Glu Val Val Asn Tyr Asp Phe Val |     |     |
| 335   | 340 | 345 |
| Ser Phe Pro Thr Asp Met Trp Ser Val Gly Val Ile Ala Tyr Met |     |     |
| 350   | 355 | 360 |
| Leu Leu Ser Gly Leu Ser Pro Phe Leu Gly Asp Asn Asp Ala Glu |     |     |
| 365   | 370 | 375 |
| Thr Leu Asn Asn Ile Leu Ala Cys Arg Trp Asp Leu Glu Asp Glu |     |     |
| 380   | 385 | 390 |
| Glu Phe Gln Asp Ile Ser Glu Glu Ala Lys Glu Phe Ile Ser Lys |     |     |
| 395   | 400 | 405 |
| Leu Leu Ile Lys Glu Lys Ser Trp Arg Ile Ser Ala Ser Glu Ala |     |     |
| 410   | 415 | 420 |
| Leu Lys His Pro Trp Leu Ser Asp His Lys Leu His Ser Arg Leu |     |     |
| 425   | 430 | 435 |
| Asn Ala Gln Val Thr Thr Ala Ser Cys Ser Ser Ser Phe Ser Pro |     |     |
| 440   | 445 | 450 |
| Val Cys Leu Ser Phe Glu Asp Gln Met Leu Glu Ser Ser         |     |     |
| 455   | 460 |     |

&lt;210&gt; 6

&lt;211&gt; 565

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7472343CD1

&lt;400&gt; 6

|   |     |     |
|---|-----|-----|
| Met Ser Thr Ile Gln Ser Glu Thr Asp Cys Tyr Asp Ile Ile Glu |     |     |
| 1   | 5   | 10  |
| Val Leu Gly Lys Gly Thr Phe Gly Glu Val Ala Lys Gly Trp Arg |     |     |
| 20  | 25  | 30  |
| Arg Ser Thr Gly Glu Met Val Ala Ile Lys Ile Leu Lys Asn Asp |     |     |
| 35  | 40  | 45  |
| Ala Tyr Arg Asn Arg Ile Ile Lys Asn Glu Leu Lys Leu Leu His |     |     |
| 50  | 55  | 60  |
| Cys Met Arg Gly Leu Asp Pro Glu Glu Ala His Val Ile Arg Phe |     |     |
| 65  | 70  | 75  |
| Leu Glu Phe Phe His Asp Ala Leu Lys Phe Tyr Leu Val Phe Glu |     |     |
| 80  | 85  | 90  |
| Leu Leu Glu Gln Asn Leu Phe Glu Phe Gln Lys Glu Asn Asn Phe |     |     |
| 95  | 100 | 105 |
| Ala Pro Leu Pro Ala Arg His Ile Arg Thr Val Thr Leu Gln Val |     |     |
| 110   | 115 | 120 |

|   |                         |
|---|-------------------------|
| Leu Thr Ala Leu Ala Arg Leu Lys Glu                         | Leu Ala Ile Ile His Ala |
| 125   | 130 135                 |
| Asp Leu Lys Pro Glu Asn Ile Met Leu Val Asp Gln Thr Arg Cys |                         |
| 140   | 145 150                 |
| Pro Phe Arg Val Lys Val Ile Asp Phe Gly Ser Ala Ser Ile Phe |                         |
| 155   | 160 165                 |
| Ser Glu Val Arg Tyr Val Lys Glu Pro Tyr Ile Gln Ser Arg Phe |                         |
| 170   | 175 180                 |
| Tyr Arg Ala Pro Glu Ile Leu Leu Gly Leu Pro Phe Cys Glu Lys |                         |
| 185   | 190 195                 |
| Val Asp Val Trp Ser Leu Gly Cys Val Met Ala Glu Leu His Leu |                         |
| 200   | 205 210                 |
| Gly Trp Pro Leu Tyr Pro Gly Asn Asn Glu Tyr Asp Gln Val Arg |                         |
| 215   | 220 225                 |
| Tyr Ile Cys Glu Thr Gln Gly Leu Pro Lys Pro His Leu Leu His |                         |
| 230   | 235 240                 |
| Ala Ala Cys Lys Ala His His Phe Phe Lys Arg Asn Pro His Pro |                         |
| 245   | 250 255                 |
| Asp Ala Ala Asn Pro Trp Gln Leu Lys Ser Ser Ala Asp Tyr Leu |                         |
| 260   | 265 270                 |
| Ala Glu Thr Lys Val Arg Glu Lys Glu Arg Arg Lys Tyr Met Leu |                         |
| 275   | 280 285                 |
| Lys Ser Leu Asp Gln Ile Glu Thr Val Asn Gly Gly Ser Val Ala |                         |
| 290   | 295 300                 |
| Ser Arg Leu Thr Phe Pro Asp Arg Glu Ala Leu Ala Glu His Ala |                         |
| 305   | 310 315                 |
| Asp Leu Lys Ser Met Val Glu Leu Ile Lys Arg Met Leu Thr Trp |                         |
| 320   | 325 330                 |
| Glu Ser His Glu Arg Ile Ser Pro Ser Ala Ala Leu Arg His Pro |                         |
| 335   | 340 345                 |
| Phe Val Ser Met Gln Gln Leu Arg Ser Ala His Glu Thr Thr His |                         |
| 350   | 355 360                 |
| Tyr Tyr Gln Leu Ser Leu Arg Ser Tyr Arg Leu Ser Leu Gln Val |                         |
| 365   | 370 375                 |
| Glu Gly Lys Pro Pro Thr Pro Val Val Ala Ala Glu Asp Gly Thr |                         |
| 380   | 385 390                 |
| Pro Tyr Tyr Cys Leu Ala Glu Glu Lys Glu Ala Ala Gly Met Gly |                         |
| 395   | 400 405                 |
| Ser Val Ala Gly Ser Ser Pro Phe Phe Arg Glu Glu Lys Ala Pro |                         |
| 410   | 415 420                 |
| Gly Met Gln Arg Ala Ile Asp Gln Leu Asp Asp Leu Ser Leu Gln |                         |
| 425   | 430 435                 |
| Glu Ala Gly His Gly Leu Trp Gly Glu Thr Cys Thr Asn Ala Val |                         |
| 440   | 445 450                 |
| Ser Asp Met Met Val Pro Leu Lys Ala Ala Ile Thr Gly His His |                         |
| 455   | 460 465                 |
| Val Pro Asp Ser Gly Pro Glu Pro Ile Leu Ala Phe Tyr Ser Ser |                         |
| 470   | 475 480                 |
| Arg Leu Ala Gly Arg His Lys Ala Arg Lys Pro Pro Ala Gly Ser |                         |
| 485   | 490 495                 |
| Lys Ser Asp Ser Asn Phe Ser Asn Leu Ile Arg Leu Ser Gln Val |                         |
| 500   | 505 510                 |
| Ser Pro Glu Asp Asp Arg Pro Cys Arg Gly Ser Ser Trp Glu Glu |                         |
| 515   | 520 525                 |
| Gly Glu His Leu Gly Ala Ser Ala Glu Pro Leu Ala Ile Leu Gln |                         |
| 530   | 535 540                 |

Arg Asp Glu Asp Gly Pro Asn Ile Asp Asn Met Thr Met Glu Ala  
                                   545                                  550                                  555  
 Glu Val Ser Arg Val Arg Ser Gly Tyr Asp  
                                   560                                  565

<210> 7  
 <211> 1319  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7480783CD1

<400> 7  
 Met Arg Leu Gly Gly Ala Arg Ala Thr Arg Arg Arg Gln Leu Leu  
   1                                  5                                  10                                  15  
 Arg Ser Ser Gly Ala Ala Gly Gly Ala Glu Leu Ala Ser Arg Arg  
                                   20                                  25                                  30  
 Arg Gly Gly Ala Gly Gly Pro Arg Gly Ala Gly Pro Pro Gly Cys  
                                   35                                  40                                  45  
 Ser Arg Ala Pro Pro Arg Leu Arg Thr Pro Ser Arg Gly Pro Gly  
                                   50                                  55                                  60  
 Val Ser Val Asn Pro Gly Ser Pro Met Gly Glu Val Glu Pro Gly  
                                   65                                  70                                  75  
 Pro Ala Gly Pro Leu Glu Pro Pro Glu Pro Pro Glu Ala Pro Ala  
                                   80                                  85                                  90  
 Ser Arg Arg Pro Gly Gly Ile Arg Val Leu Lys Ile Val Tyr Asp  
                                   95                                  100                                  105  
 Tyr Leu Ser Arg Leu Gly Phe Asp Asp Pro Val Arg Ile Gln Glu  
                                   110                                  115                                  120  
 Glu Ala Thr Asn Pro Asp Leu Gly Cys Met Ile Arg Phe Tyr Gly  
                                   125                                  130                                  135  
 Glu Lys Pro Cys His Met Asp Arg Leu Asp Arg Ile Leu Leu Ser  
                                   140                                  145                                  150  
 Gly Ile Tyr Asn Val Arg Lys Gly Lys Thr Gln Leu His Lys Trp  
                                   155                                  160                                  165  
 Ala Glu Arg Leu Val Val Leu Cys Gly Thr Cys Leu Ile Val Ser  
                                   170                                  175                                  180  
 Ser Val Lys Asp Cys Gln Thr Gly Lys Met His Ile Leu Pro Leu  
                                   185                                  190                                  195  
 Val Gly Gly Lys Ile Glu Glu Val Lys Arg Arg Gln Tyr Ser Leu  
                                   200                                  205                                  210  
 Ala Phe Ser Ser Ala Gly Ala Gln Ala Gln Thr Tyr His Val Ser  
                                   215                                  220                                  225  
 Phe Glu Thr Leu Ala Glu Tyr Gln Arg Trp Gln Arg Gln Ala Ser  
                                   230                                  235                                  240  
 Lys Val Val Ser Gln Arg Ile Ser Thr Val Asp Leu Ser Cys Tyr  
                                   245                                  250                                  255  
 Ser Leu Glu Glu Val Pro Glu His Leu Phe Tyr Ser Gln Asp Ile  
                                   260                                  265                                  270  
 Thr Tyr Leu Asn Leu Arg His Asn Phe Met Gln Leu Glu Arg Pro  
                                   275                                  280                                  285  
 Gly Gly Leu Asp Thr Leu Tyr Lys Phe Ser Gln Leu Lys Gly Leu  
                                   290                                  295                                  300  
 Asn Leu Ser His Asn Lys Leu Gly Leu Phe Pro Ile Leu Leu Cys

|   |     |     |
|---|-----|-----|
| 305   | 310 | 315 |
| Glu Ile Ser Thr Leu Thr Glu Leu Asn Leu Ser Cys Asn Gly Phe |     |     |
| 320   | 325 | 330 |
| His Asp Leu Pro Ser Gln Ile Gly Asn Leu Leu Asn Leu Gln Thr |     |     |
| 335   | 340 | 345 |
| Leu Cys Leu Asp Gly Asn Phe Leu Thr Thr Leu Pro Glu Glu Leu |     |     |
| 350   | 355 | 360 |
| Gly Asn Leu Gln Gln Leu Ser Ser Leu Gly Ile Ser Phe Asn Asn |     |     |
| 365   | 370 | 375 |
| Phe Ser Gln Ile Pro Glu Val Tyr Glu Lys Leu Thr Met Leu Asp |     |     |
| 380   | 385 | 390 |
| Arg Val Val Met Ala Gly Asn Cys Leu Glu Val Leu Asn Leu Gly |     |     |
| 395   | 400 | 405 |
| Val Leu Asn Arg Met Asn His Ile Lys His Val Asp Leu Arg Met |     |     |
| 410   | 415 | 420 |
| Asn His Leu Lys Thr Met Val Ile Glu Asn Leu Glu Gly Asn Lys |     |     |
| 425   | 430 | 435 |
| His Ile Thr His Val Asp Leu Arg Asp Asn Arg Leu Thr Asp Leu |     |     |
| 440   | 445 | 450 |
| Asp Leu Ser Ser Leu Cys Ser Leu Glu Gln Leu His Cys Gly Arg |     |     |
| 455   | 460 | 465 |
| Asn Gln Leu Arg Glu Leu Thr Leu Ser Gly Phe Ser Leu Arg Thr |     |     |
| 470   | 475 | 480 |
| Leu Tyr Ala Ser Ser Asn Arg Leu Thr Ala Val Asn Val Tyr Pro |     |     |
| 485   | 490 | 495 |
| Val Pro Ser Leu Leu Thr Phe Leu Asp Leu Ser Arg Asn Leu Leu |     |     |
| 500   | 505 | 510 |
| Glu Cys Val Pro Asp Trp Ala Cys Glu Ala Lys Lys Ile Glu Val |     |     |
| 515   | 520 | 525 |
| Leu Asp Val Ser Tyr Asn Leu Leu Thr Glu Val Pro Val Arg Ile |     |     |
| 530   | 535 | 540 |
| Leu Ser Ser Leu Ser Leu Arg Lys Leu Met Leu Gly His Asn His |     |     |
| 545   | 550 | 555 |
| Val Gln Asn Leu Pro Thr Leu Val Glu His Ile Pro Leu Glu Val |     |     |
| 560   | 565 | 570 |
| Leu Asp Leu Gln His Asn Ala Leu Thr Arg Leu Pro Asp Thr Leu |     |     |
| 575   | 580 | 585 |
| Phe Ser Lys Ala Leu Asn Leu Arg Tyr Leu Asn Ala Ser Ala Asn |     |     |
| 590   | 595 | 600 |
| Ser Leu Glu Ser Leu Pro Ser Ala Cys Thr Gly Glu Glu Ser Leu |     |     |
| 605   | 610 | 615 |
| Ser Met Leu Gln Leu Leu Tyr Leu Thr Asn Asn Leu Leu Thr Asp |     |     |
| 620   | 625 | 630 |
| Gln Cys Ile Pro Val Leu Val Gly His Leu His Leu Arg Ile Leu |     |     |
| 635   | 640 | 645 |
| His Leu Ala Asn Asn Gln Leu Gln Thr Phe Pro Ala Ser Lys Leu |     |     |
| 650   | 655 | 660 |
| Asn Lys Leu Glu Gln Leu Glu Glu Leu Asn Leu Ser Gly Asn Lys |     |     |
| 665   | 670 | 675 |
| Leu Lys Thr Ile Pro Thr Thr Ile Ala Asn Cys Lys Arg Leu His |     |     |
| 680   | 685 | 690 |
| Thr Leu Val Ala His Ser Asn Asn Ile Ser Ile Phe Pro Glu Ile |     |     |
| 695   | 700 | 705 |
| Leu Gln Leu Pro Gln Ile Gln Phe Val Asp Leu Ser Cys Asn Asp |     |     |
| 710   | 715 | 720 |
| Leu Thr Glu Ile Leu Ile Pro Glu Ala Leu Pro Ala Thr Leu Gln |     |     |

|   |      |  |      |  |      |
|---|------|--|------|--|------|
|   | 725  |  | 730  |  | 735  |
| Asp Leu Asp Leu Thr Gly Asn Thr Asn Leu Val Leu Glu His Lys |      |  |      |  |      |
|   | 740  |  | 745  |  | 750  |
| Thr Leu Asp Ile Phe Ser His Ile Thr Thr Leu Lys Ile Asp Gln |      |  |      |  |      |
|   | 755  |  | 760  |  | 765  |
| Lys Pro Leu Pro Thr Thr Asp Ser Thr Val Thr Ser Thr Phe Trp |      |  |      |  |      |
|   | 770  |  | 775  |  | 780  |
| Ser His Gly Leu Ala Glu Met Ala Gly Gln Arg Asn Lys Leu Cys |      |  |      |  |      |
|   | 785  |  | 790  |  | 795  |
| Val Ser Ala Leu Ala Met Asp Ser Phe Ala Glu Gly Val Gly Ala |      |  |      |  |      |
|   | 800  |  | 805  |  | 810  |
| Val Tyr Gly Met Phe Asp Gly Asp Arg Asn Glu Glu Leu Pro Arg |      |  |      |  |      |
|   | 815  |  | 820  |  | 825  |
| Leu Leu Gln Cys Thr Met Ala Asp Val Leu Leu Glu Glu Val Gln |      |  |      |  |      |
|   | 830  |  | 835  |  | 840  |
| Gln Ser Thr Asn Asp Thr Val Phe Met Ala Asn Thr Phe Leu Val |      |  |      |  |      |
|   | 845  |  | 850  |  | 855  |
| Ser His Arg Lys Leu Gly Met Ala Gly Gln Lys Leu Gly Ser Ser |      |  |      |  |      |
|   | 860  |  | 865  |  | 870  |
| Ala Leu Leu Cys Tyr Ile Arg Pro Asp Thr Ala Asp Pro Ala Ser |      |  |      |  |      |
|   | 875  |  | 880  |  | 885  |
| Ser Phe Ser Leu Thr Val Ala Asn Val Gly Thr Cys Gln Ala Val |      |  |      |  |      |
|   | 890  |  | 895  |  | 900  |
| Leu Cys Arg Gly Gly Lys Pro Val Pro Leu Ser Lys Val Phe Ser |      |  |      |  |      |
|   | 905  |  | 910  |  | 915  |
| Leu Glu Gln Asp Pro Glu Glu Ala Gln Arg Val Lys Asp Gln Lys |      |  |      |  |      |
|   | 920  |  | 925  |  | 930  |
| Ala Ile Ile Thr Glu Asp Asn Lys Val Asn Gly Val Thr Cys Cys |      |  |      |  |      |
|   | 935  |  | 940  |  | 945  |
| Thr Arg Met Leu Gly Cys Thr Tyr Leu Tyr Pro Trp Ile Leu Pro |      |  |      |  |      |
|   | 950  |  | 955  |  | 960  |
| Lys Pro His Ile Ser Ser Thr Pro Leu Thr Ile Gln Asp Glu Leu |      |  |      |  |      |
|   | 965  |  | 970  |  | 975  |
| Leu Ile Leu Gly Asn Lys Ala Leu Trp Glu His Ile Ser Tyr Thr |      |  |      |  |      |
|   | 980  |  | 985  |  | 990  |
| Glu Ala Val Asn Ala Val Arg His Val Gln Asp Pro Leu Ala Ala |      |  |      |  |      |
|   | 995  |  | 1000 |  | 1005 |
| Ala Lys Lys Leu Cys Thr Leu Ala Gln Ser Tyr Gly Cys Gln Asp |      |  |      |  |      |
|   | 1010 |  | 1015 |  | 1020 |
| Asn Val Gly Ala Met Val Val Tyr Leu Asn Ile Gly Glu Glu Gly |      |  |      |  |      |
|   | 1025 |  | 1030 |  | 1035 |
| Cys Thr Cys Glu Met Asn Gly Leu Thr Leu Pro Gly Pro Val Gly |      |  |      |  |      |
|   | 1040 |  | 1045 |  | 1050 |
| Phe Ala Ser Thr Thr Thr Ile Lys Asp Ala Pro Lys Pro Ala Thr |      |  |      |  |      |
|   | 1055 |  | 1060 |  | 1065 |
| Pro Ser Ser Ser Ser Gly Ile Ala Ser Glu Phe Ser Ser Glu Met |      |  |      |  |      |
|   | 1070 |  | 1075 |  | 1080 |
| Ser Thr Ser Glu Val Ser Ser Glu Val Gly Ser Thr Ala Ser Asp |      |  |      |  |      |
|   | 1085 |  | 1090 |  | 1095 |
| Glu His Asn Ala Gly Gly Leu Asp Thr Ala Leu Leu Pro Arg Pro |      |  |      |  |      |
|   | 1100 |  | 1105 |  | 1110 |
| Glu Arg Arg Cys Ser Leu His Pro Thr Pro Thr Ser Gly Leu Phe |      |  |      |  |      |
|   | 1115 |  | 1120 |  | 1125 |
| Gln Arg Gln Pro Ser Ser Ala Thr Phe Ser Ser Asn Gln Ser Asp |      |  |      |  |      |
|   | 1130 |  | 1135 |  | 1140 |
| Asn Gly Leu Asp Ser Asp Asp Asp Gln Pro Val Glu Gly Val Ile |      |  |      |  |      |

|   |      |      |
|---|------|------|
| 1145  | 1150 | 1155 |
| Thr Asn Gly Ser Lys Val Glu Val Glu Val Asp Ile His Cys Cys |      |      |
| 1160  | 1165 | 1170 |
| Arg Gly Arg Asp Leu Glu Asn Ser Pro Pro Leu Ile Glu Ser Ser |      |      |
| 1175  | 1180 | 1185 |
| Pro Thr Leu Cys Ser Glu Glu His Ala Arg Gly Ser Cys Phe Gly |      |      |
| 1190  | 1195 | 1200 |
| Ile Arg Arg Gln Asn Ser Val Asn Ser Gly Met Leu Leu Pro Met |      |      |
| 1205  | 1210 | 1215 |
| Ser Lys Asp Arg Met Glu Leu Gln Lys Ser Pro Ser Thr Ser Cys |      |      |
| 1220  | 1225 | 1230 |
| Leu Tyr Gly Lys Lys Leu Ser Asn Gly Ser Ile Val Pro Leu Glu |      |      |
| 1235  | 1240 | 1245 |
| Asp Ser Leu Asn Leu Ile Glu Val Ala Thr Glu Val Pro Lys Arg |      |      |
| 1250  | 1255 | 1260 |
| Lys Thr Gly Tyr Phe Ala Ala Pro Thr Gln Met Glu Pro Glu Asp |      |      |
| 1265  | 1270 | 1275 |
| Gln Phe Val Val Pro His Asp Leu Glu Glu Glu Val Lys Glu Gln |      |      |
| 1280  | 1285 | 1290 |
| Met Lys Gln His Gln Asp Ser Arg Leu Glu Pro Glu Pro His Glu |      |      |
| 1295  | 1300 | 1305 |
| Glu Asp Arg Thr Glu Pro Pro Glu Glu Phe Asp Thr Ala Leu     |      |      |
| 1310  | 1315 |      |

&lt;210&gt; 8

&lt;211&gt; 414

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7477063CD1

&lt;400&gt; 8

|   |     |     |
|---|-----|-----|
| Met Asp Ser Glu Thr His Ser Gly Met His Arg Gly Arg Gly Arg |     |     |
| 1   | 5   | 10  |
| Trp Arg Glu Arg Pro Gly Trp Ala Gly Gly Leu Cys Gly Leu Arg |     |     |
| 20  | 25  | 30  |
| Met His Pro His Ser Gly Leu Gly Ala Pro Gly Leu Leu Pro Gln |     |     |
| 35  | 40  | 45  |
| Thr Gly Ala Gly Gly Ala Ser Val Ala Val Thr Pro Asn Leu Ser |     |     |
| 50  | 55  | 60  |
| Arg Thr Gln Lys Gln Val Ala Arg Val Arg Glu Asp Thr Ala Thr |     |     |
| 65  | 70  | 75  |
| Ala Leu Gln Arg Leu Val Glu Leu Thr Thr Ser Arg Val Thr Pro |     |     |
| 80  | 85  | 90  |
| Val Arg Ser Leu Arg Asp Gln Tyr His Leu Ile Arg Lys Leu Gly |     |     |
| 95  | 100 | 105 |
| Ser Gly Ser Tyr Gly Arg Val Leu Leu Ala Gln Pro His Gln Gly |     |     |
| 110   | 115 | 120 |
| Gly Pro Ala Val Ala Leu Lys Leu Leu Arg Arg Asp Leu Val Leu |     |     |
| 125   | 130 | 135 |
| Arg Ser Thr Phe Leu Arg Glu Phe Cys Val Gly Arg Cys Val Ser |     |     |
| 140   | 145 | 150 |
| Ala His Pro Gly Leu Leu Gln Thr Leu Ala Gly Pro Leu Gln Thr |     |     |
| 155   | 160 | 165 |

```

Pro Arg Tyr Phe Ala Phe Ala Gln Glu Tyr Ala Pro Cys Gly Asp
      170                      175                      180
Leu Ser Gly Met Leu Gln Glu Arg Gly Leu Pro Glu Leu Leu Val
      185                      190                      195
Lys Arg Val Val Ala Gln Leu Ala Gly Ala Leu Asp Phe Leu His
      200                      205                      210
Ser Arg Gly Leu Val His Ala Asp Val Lys Pro Asp Asn Val Leu
      215                      220                      225
Val Phe Asp Pro Val Cys Ser Arg Val Ala Leu Gly Asp Leu Gly
      230                      235                      240
Leu Thr Arg Pro Glu Gly Ser Pro Thr Pro Ala Pro Pro Val Pro
      245                      250                      255
Leu Pro Thr Ala Pro Pro Glu Leu Cys Leu Leu Leu Pro Pro Asp
      260                      265                      270
Thr Leu Pro Leu Arg Pro Ala Val Asp Ser Trp Gly Leu Gly Val
      275                      280                      285
Leu Leu Phe Cys Ala Ala Thr Ala Cys Phe Pro Trp Asp Val Ala
      290                      295                      300
Leu Ala Pro Asn Pro Glu Phe Glu Ala Phe Ala Gly Trp Val Thr
      305                      310                      315
Thr Lys Pro Gln Pro Pro Gln Pro Pro Pro Pro Trp Asp Gln Phe
      320                      325                      330
Ala Pro Pro Ala Leu Ala Leu Leu Gln Gly Leu Leu Asp Leu Asp
      335                      340                      345
Pro Glu Thr Arg Ser Pro Pro Leu Ala Val Leu Asp Phe Leu Gly
      350                      355                      360
Asp Asp Trp Gly Leu Gln Gly Asn Arg Glu Gly Pro Gly Val Leu
      365                      370                      375
Gly Ser Ala Val Ser Tyr Glu Asp Arg Glu Gly Gly Ser Ser
      380                      385                      390
Leu Glu Glu Trp Thr Asp Glu Gly Asp Asp Ser Lys Ser Gly Gly
      395                      400                      405
Arg Thr Gly Thr Asp Gly Gly Ala Pro
      410

```

<210> 9  
 <211> 1036  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7475394CD1

```

<400> 9
Met Ala Leu Arg Gly Ala Ala Gly Ala Thr Asp Thr Pro Val Ser
  1                      5                      10                      15
Ser Ala Gly Gly Ala Pro Gly Gly Ser Ala Ser Ser Ser Ser Thr
      20                      25                      30
Ser Ser Gly Gly Ser Ala Ser Ala Gly Ala Gly Leu Trp Ala Ala
      35                      40                      45
Leu Tyr Asp Tyr Glu Ala Arg Gly Glu Asp Glu Leu Ser Leu Arg
      50                      55                      60
Arg Gly Gln Leu Val Glu Val Leu Ser Gln Asp Ala Ala Val Ser
      65                      70                      75
Gly Asp Glu Gly Trp Trp Ala Gly Gln Val Gln Arg Arg Leu Gly

```

|   |     |  |     |  |     |
|---|-----|--|-----|--|-----|
|   | 80  |  | 85  |  | 90  |
| Ile Phe Pro Ala Asn Tyr Val Ala Pro Cys Arg Pro Ala Ala Ser |     |  |     |  |     |
|   | 95  |  | 100 |  | 105 |
| Pro Ala Pro Pro Pro Ser Arg Pro Ser Ser Pro Val His Val Ala |     |  |     |  |     |
|   | 110 |  | 115 |  | 120 |
| Phe Glu Arg Leu Glu Leu Lys Glu Leu Ile Gly Ala Gly Gly Phe |     |  |     |  |     |
|   | 125 |  | 130 |  | 135 |
| Gly Gln Val Tyr Arg Ala Thr Trp Gln Gly Gln Glu Val Ala Val |     |  |     |  |     |
|   | 140 |  | 145 |  | 150 |
| Lys Ala Ala Arg Gln Asp Pro Glu Gln Asp Ala Ala Ala Ala Ala |     |  |     |  |     |
|   | 155 |  | 160 |  | 165 |
| Glu Ser Val Arg Arg Glu Ala Arg Leu Phe Ala Met Leu Arg His |     |  |     |  |     |
|   | 170 |  | 175 |  | 180 |
| Pro Asn Ile Ile Glu Leu Arg Gly Val Cys Leu Gln Gln Pro His |     |  |     |  |     |
|   | 185 |  | 190 |  | 195 |
| Leu Cys Leu Val Leu Glu Phe Ala Arg Gly Gly Ala Leu Asn Arg |     |  |     |  |     |
|   | 200 |  | 205 |  | 210 |
| Ala Leu Ala Ala Ala Asn Ala Ala Pro Asp Pro Arg Ala Pro Gly |     |  |     |  |     |
|   | 215 |  | 220 |  | 225 |
| Pro Arg Arg Ala Arg Arg Ile Pro Pro His Val Leu Val Asn Trp |     |  |     |  |     |
|   | 230 |  | 235 |  | 240 |
| Ala Val Gln Ile Ala Arg Gly Met Leu Tyr Leu His Glu Glu Ala |     |  |     |  |     |
|   | 245 |  | 250 |  | 255 |
| Phe Val Pro Ile Leu His Arg Asp Leu Lys Ser Ser Asn Ile Leu |     |  |     |  |     |
|   | 260 |  | 265 |  | 270 |
| Leu Leu Glu Lys Ile Glu His Asp Asp Ile Cys Asn Lys Thr Leu |     |  |     |  |     |
|   | 275 |  | 280 |  | 285 |
| Lys Ile Thr Asp Phe Gly Leu Ala Arg Glu Trp His Arg Thr Thr |     |  |     |  |     |
|   | 290 |  | 295 |  | 300 |
| Lys Met Ser Thr Ala Gly Thr Tyr Ala Trp Met Ala Pro Glu Val |     |  |     |  |     |
|   | 305 |  | 310 |  | 315 |
| Ile Lys Ser Ser Leu Phe Ser Lys Gly Ser Asp Ile Trp Ser Tyr |     |  |     |  |     |
|   | 320 |  | 325 |  | 330 |
| Gly Val Leu Leu Trp Glu Leu Leu Thr Gly Glu Val Pro Tyr Arg |     |  |     |  |     |
|   | 335 |  | 340 |  | 345 |
| Gly Ile Asp Gly Leu Ala Val Ala Tyr Gly Val Ala Val Asn Lys |     |  |     |  |     |
|   | 350 |  | 355 |  | 360 |
| Leu Thr Leu Pro Ile Pro Ser Thr Cys Pro Glu Pro Phe Ala Lys |     |  |     |  |     |
|   | 365 |  | 370 |  | 375 |
| Leu Met Lys Glu Cys Trp Gln Gln Asp Pro His Ile Arg Pro Ser |     |  |     |  |     |
|   | 380 |  | 385 |  | 390 |
| Phe Ala Leu Ile Leu Glu Gln Leu Thr Ala Ile Glu Gly Ala Val |     |  |     |  |     |
|   | 395 |  | 400 |  | 405 |
| Met Thr Glu Met Pro Gln Glu Ser Phe His Ser Met Gln Asp Asp |     |  |     |  |     |
|   | 410 |  | 415 |  | 420 |
| Trp Lys Leu Glu Ile Gln Gln Met Phe Asp Glu Leu Arg Thr Lys |     |  |     |  |     |
|   | 425 |  | 430 |  | 435 |
| Glu Lys Glu Leu Arg Ser Arg Glu Glu Glu Leu Thr Arg Ala Ala |     |  |     |  |     |
|   | 440 |  | 445 |  | 450 |
| Leu Gln Gln Lys Ser Gln Glu Glu Leu Leu Lys Arg Arg Glu Gln |     |  |     |  |     |
|   | 455 |  | 460 |  | 465 |
| Gln Leu Ala Glu Arg Glu Ile Asp Val Leu Glu Arg Glu Leu Asn |     |  |     |  |     |
|   | 470 |  | 475 |  | 480 |
| Ile Leu Ile Phe Gln Leu Asn Gln Glu Lys Pro Lys Val Lys Lys |     |  |     |  |     |
|   | 485 |  | 490 |  | 495 |
| Arg Lys Gly Lys Phe Lys Arg Ser Arg Leu Lys Leu Lys Asp Gly |     |  |     |  |     |



|                                     |                         |     |
|-------------------------------------|-------------------------|-----|
| 500                                 | 505                     | 510 |
| His Arg Ile Ser Leu Pro Ser Asp Phe | Gln His Lys Ile Thr Val |     |
| 515                                 | 520                     | 525 |
| Gln Ala Ser Pro Asn Leu Asp Lys Arg | Arg Ser Leu Asn Ser Ser |     |
| 530                                 | 535                     | 540 |
| Ser Ser Ser Pro Pro Ser Ser Pro Thr | Met Met Pro Arg Leu Arg |     |
| 545                                 | 550                     | 555 |
| Ala Ile Gln Leu Thr Ser Asp Glu Ser | Asn Lys Thr Trp Gly Arg |     |
| 560                                 | 565                     | 570 |
| Asn Thr Val Phe Arg Gln Glu Glu Phe | Glu Asp Val Lys Arg Asn |     |
| 575                                 | 580                     | 585 |
| Phe Lys Lys Lys Gly Cys Thr Trp Gly | Pro Asn Ser Ile Gln Met |     |
| 590                                 | 595                     | 600 |
| Lys Asp Arg Thr Asp Cys Lys Glu Arg | Ile Arg Pro Leu Ser Asp |     |
| 605                                 | 610                     | 615 |
| Gly Asn Ser Pro Trp Ser Thr Ile Leu | Ile Lys Asn Gln Lys Thr |     |
| 620                                 | 625                     | 630 |
| Met Pro Leu Ala Ser Leu Phe Val Asp | Gln Pro Gly Ser Cys Glu |     |
| 635                                 | 640                     | 645 |
| Glu Pro Lys Leu Ser Pro Asp Gly Leu | Glu His Arg Lys Pro Lys |     |
| 650                                 | 655                     | 660 |
| Gln Ile Lys Leu Pro Ser Gln Ala Tyr | Ile Asp Leu Pro Leu Gly |     |
| 665                                 | 670                     | 675 |
| Lys Asp Ala Gln Arg Glu Asn Pro Ala | Glu Ala Glu Ser Trp Glu |     |
| 680                                 | 685                     | 690 |
| Glu Ala Ala Ser Ala Asn Ala Ala Thr | Val Ser Ile Glu Met Thr |     |
| 695                                 | 700                     | 705 |
| Pro Thr Asn Ser Leu Ser Arg Ser Pro | Gln Arg Lys Lys Thr Glu |     |
| 710                                 | 715                     | 720 |
| Ser Ala Leu Tyr Gly Cys Thr Val Leu | Leu Ala Ser Val Ala Leu |     |
| 725                                 | 730                     | 735 |
| Gly Leu Asp Leu Arg Glu Leu His Lys | Ala Gln Ala Ala Glu Glu |     |
| 740                                 | 745                     | 750 |
| Pro Leu Pro Lys Glu Glu Lys Lys Lys | Arg Glu Gly Ile Phe Gln |     |
| 755                                 | 760                     | 765 |
| Arg Ala Ser Lys Ser Arg Arg Ser Ala | Ser Pro Pro Thr Ser Leu |     |
| 770                                 | 775                     | 780 |
| Pro Ser Thr Cys Gly Glu Ala Ser Ser | Pro Pro Ser Leu Pro Leu |     |
| 785                                 | 790                     | 795 |
| Ser Ser Ala Leu Gly Ile Leu Ser Thr | Pro Ser Phe Ser Thr Lys |     |
| 800                                 | 805                     | 810 |
| Cys Leu Leu Gln Met Asp Ser Glu Asp | Pro Leu Val Asp Ser Ala |     |
| 815                                 | 820                     | 825 |
| Pro Val Thr Cys Asp Ser Glu Met Leu | Thr Pro Asp Phe Cys Pro |     |
| 830                                 | 835                     | 840 |
| Thr Ala Pro Gly Ser Gly Arg Glu Pro | Ala Leu Met Pro Arg Leu |     |
| 845                                 | 850                     | 855 |
| Asp Thr Asp Cys Ser Val Ser Arg Asn | Leu Pro Ser Ser Phe Leu |     |
| 860                                 | 865                     | 870 |
| Gln Gln Thr Cys Gly Asn Val Pro Tyr | Cys Ala Ser Ser Lys His |     |
| 875                                 | 880                     | 885 |
| Arg Pro Ser His His Arg Arg Thr Met | Ser Asp Gly Asn Pro Thr |     |
| 890                                 | 895                     | 900 |
| Pro Thr Gly Ala Thr Ile Ile Ser Ala | Thr Gly Ala Ser Ala Leu |     |
| 905                                 | 910                     | 915 |
| Pro Leu Cys Pro Ser Pro Ala Pro His | Ser His Leu Pro Arg Glu |     |

|                 |   |                         |      |
|-----------------|---|-------------------------|------|
|                 | 920   | 925                     | 930  |
| Val Ser Pro Lys | Lys His Ser Thr Val                         | His Ile Val Pro Gln Arg |      |
|                 | 935   | 940                     | 945  |
| Arg Pro Ala Ser | Leu Arg Ser Arg Ser Asp Leu Pro Gln Ala Tyr |                         |      |
|                 | 950   | 955                     | 960  |
| Pro Gln Thr Ala | Val Ser Gln Leu Ala Gln Thr Ala Cys Val Val |                         |      |
|                 | 965   | 970                     | 975  |
| Gly Arg Pro Gly | Pro His Pro Thr Gln Phe Leu Ala Ala Lys Glu |                         |      |
|                 | 980   | 985                     | 990  |
| Arg Thr Lys Ser | His Val Pro Ser Leu Leu Asp Ala Asp Val Glu |                         |      |
|                 | 995   | 1000                    | 1005 |
| Gly Gln Ser Arg | Asp Tyr Thr Val Pro Leu Cys Arg Met Arg Ser |                         |      |
|                 | 1010  | 1015                    | 1020 |
| Lys Thr Ser Arg | Pro Ser Ile Tyr Glu Leu Glu Lys Glu Phe Leu |                         |      |
|                 | 1025  | 1030                    | 1035 |
| Ser             |   |                         |      |

&lt;210&gt; 10

&lt;211&gt; 293

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7482884CD1

&lt;400&gt; 10

|   |     |     |
|---|-----|-----|
| Met Pro Glu Asn Ser Asn Phe Pro Tyr Arg Arg Tyr Asp Arg Leu |     |     |
| 1   | 5   | 10  |
| Pro Pro Ile His Gln Phe Ser Ile Glu Ser Asp Thr Asp Leu Ser |     |     |
|   | 20  | 25  |
| Glu Thr Ala Glu Leu Ile Glu Glu Tyr Glu Val Phe Asp Pro Thr |     |     |
|   | 35  | 40  |
| Arg Pro Arg Pro Lys Ile Ile Leu Val Ile Gly Gly Pro Gly Ser |     |     |
|   | 50  | 55  |
| Gly Lys Gly Thr Gln Ser Leu Lys Ile Ala Glu Arg Tyr Gly Phe |     |     |
|   | 65  | 70  |
| Gln Tyr Ile Ser Val Gly Glu Leu Leu Arg Lys Lys Ile His Ser |     |     |
|   | 80  | 85  |
| Thr Ser Ser Asn Arg Lys Trp Ser Leu Ile Ala Lys Ile Ile Thr |     |     |
|   | 95  | 100 |
| Thr Gly Glu Leu Ala Pro Gln Glu Thr Thr Ile Thr Glu Ile Lys |     |     |
|   | 110 | 115 |
| Gln Lys Leu Met Gln Ile Pro Asp Glu Glu Gly Ile Val Ile Asp |     |     |
|   | 125 | 130 |
| Gly Phe Pro Arg Asp Val Ala Gln Ala Leu Ser Phe Glu Asp Gln |     |     |
|   | 140 | 145 |
| Ile Cys Thr Pro Asp Leu Val Val Phe Leu Ala Cys Ala Asn Gln |     |     |
|   | 155 | 160 |
| Arg Leu Lys Glu Arg Leu Leu Lys Arg Ala Glu Gln Gln Gly Arg |     |     |
|   | 170 | 175 |
| Pro Asp Asp Asn Val Lys Ala Thr Gln Arg Arg Leu Met Asn Phe |     |     |
|   | 185 | 190 |
| Lys Gln Asn Ala Ala Pro Leu Val Lys Tyr Phe Gln Glu Lys Gly |     |     |
|   | 200 | 205 |
|   |     | 210 |

Leu Ile Met Thr Phe Asp Ala Asp Arg Asp Glu Asp Glu Val Phe  
                           215                          220                          225  
 Tyr Asp Ile Ser Met Ala Val Asp Asn Lys Leu Phe Pro Asn Lys  
                           230                          235                          240  
 Glu Ala Ala Ala Gly Ser Ser Asp Leu Asp Pro Ser Met Ile Leu  
                           245                          250                          255  
 Asp Thr Gly Glu Ile Ile Asp Thr Gly Ser Asp Tyr Glu Asp Gln  
                           260                          265                          270  
 Gly Asp Asp Gln Leu Asn Val Phe Gly Glu Asp Thr Met Gly Gly  
                           275                          280                          285  
 Phe Met Glu Asp Leu Arg Lys Val  
                           290

&lt;210&gt; 11

&lt;211&gt; 550

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7494121CD1

&lt;400&gt; 11

Met Asn Glu Ser Pro Asp Pro Thr Gly Leu Thr Gly Val Ile Ile  
   1                          5                          10                          15  
 Glu Leu Gly Pro Asn Asp Ser Pro Gln Thr Ser Glu Phe Lys Gly  
                           20                          25                          30  
 Ala Thr Glu Glu Ala Pro Ala Lys Glu Ser Pro His Thr Ser Glu  
                           35                          40                          45  
 Phe Lys Gly Ala Ala Arg Val Ser Pro Ile Ser Glu Ser Val Leu  
                           50                          55                          60  
 Ala Arg Leu Ser Lys Phe Glu Val Glu Asp Ala Glu Asn Val Ala  
                           65                          70                          75  
 Ser Tyr Asp Ser Lys Ile Lys Lys Ile Val His Ser Ile Val Ser  
                           80                          85                          90  
 Ser Phe Ala Phe Gly Leu Phe Gly Val Phe Leu Val Leu Leu Asp  
                           95                          100                          105  
 Val Thr Leu Ile Leu Ala Asp Leu Ile Phe Thr Asp Ser Lys Leu  
                           110                          115                          120  
 Tyr Ile Pro Leu Glu Tyr Arg Ser Ile Ser Leu Ala Ile Ala Leu  
                           125                          130                          135  
 Phe Phe Leu Met Asp Val Leu Leu Arg Val Phe Val Glu Arg Arg  
                           140                          145                          150  
 Gln Gln Tyr Phe Ser Asp Leu Phe Asn Ile Leu Asp Thr Ala Ile  
                           155                          160                          165  
 Ile Val Ile Leu Leu Leu Val Asp Val Val Tyr Ile Phe Phe Asp  
                           170                          175                          180  
 Ile Lys Leu Leu Arg Asn Ile Pro Arg Trp Thr His Leu Leu Arg  
                           185                          190                          195  
 Leu Leu Arg Leu Ile Ile Leu Leu Arg Ile Phe His Leu Phe His  
                           200                          205                          210  
 Gln Lys Arg Gln Leu Glu Lys Leu Ile Arg Arg Arg Val Ser Glu  
                           215                          220                          225  
 Asn Lys Arg Arg Tyr Thr Arg Asp Gly Phe Asp Leu Asp Leu Thr  
                           230                          235                          240  
 Tyr Val Thr Glu Arg Ile Ile Ala Met Ser Phe Pro Ser Ser Gly

|   |     |     |     |
|---|-----|-----|-----|
|   | 245 | 250 | 255 |
| Arg Gln Ser Phe Tyr Arg Asn Pro Ile Lys Glu Val Val Arg Phe |     |     |     |
|   | 260 | 265 | 270 |
| Leu Asp Lys Lys His Arg Asn His Tyr Arg Val Tyr Asn Leu Cys |     |     |     |
|   | 275 | 280 | 285 |
| Ser Glu Arg Ala Tyr Asp Pro Lys His Phe His Asn Arg Val Ser |     |     |     |
|   | 290 | 295 | 300 |
| Arg Ile Met Ile Asp Asp His Asn Val Pro Thr Leu His Gln Met |     |     |     |
|   | 305 | 310 | 315 |
| Val Val Phe Thr Lys Glu Val Asn Glu Trp Met Ala Gln Asp Leu |     |     |     |
|   | 320 | 325 | 330 |
| Glu Asn Ile Val Ala Ile His Cys Lys Gly Gly Lys Gly Arg Thr |     |     |     |
|   | 335 | 340 | 345 |
| Gly Thr Met Val Cys Ala Leu Leu Ile Ala Ser Glu Ile Phe Leu |     |     |     |
|   | 350 | 355 | 360 |
| Thr Ala Glu Glu Ser Leu Tyr Tyr Phe Gly Glu Arg Arg Thr Asn |     |     |     |
|   | 365 | 370 | 375 |
| Lys Thr His Ser Asn Lys Phe Gln Gly Val Glu Thr Pro Ser Gln |     |     |     |
|   | 380 | 385 | 390 |
| Asn Arg Tyr Val Gly Tyr Phe Ala Gln Val Lys His Leu Tyr Asn |     |     |     |
|   | 395 | 400 | 405 |
| Trp Asn Leu Pro Pro Arg Arg Ile Leu Phe Ile Lys Arg Phe Ile |     |     |     |
|   | 410 | 415 | 420 |
| Ile Tyr Ser Ile Arg Gly Asp Val Cys Asp Leu Lys Val Gln Val |     |     |     |
|   | 425 | 430 | 435 |
| Val Met Glu Lys Lys Val Val Phe Ser Ser Thr Ser Leu Gly Asn |     |     |     |
|   | 440 | 445 | 450 |
| Cys Ser Ile Leu His Asp Ile Glu Thr Asp Lys Ile Leu Ile Asn |     |     |     |
|   | 455 | 460 | 465 |
| Val Tyr Asp Gly Pro Pro Leu Tyr Asp Asp Val Lys Val Gln Phe |     |     |     |
|   | 470 | 475 | 480 |
| Phe Ser Ser Asn Leu Pro Lys Tyr Tyr Asp Asn Cys Pro Phe Phe |     |     |     |
|   | 485 | 490 | 495 |
| Phe Trp Phe Asn Thr Ser Phe Ile Gln Asn Asn Arg Leu Cys Leu |     |     |     |
|   | 500 | 505 | 510 |
| Pro Arg Asn Glu Leu Asp Asn Pro His Lys Gln Lys Ala Trp Lys |     |     |     |
|   | 515 | 520 | 525 |
| Ile Tyr Pro Pro Glu Phe Ala Val Glu Ile Leu Phe Gly Glu Met |     |     |     |
|   | 530 | 535 | 540 |
| Thr Ser Asn Asp Val Val Ala Gly Ser Asp                     |     |     |     |
|   | 545 | 550 |     |

&lt;210&gt; 12

&lt;211&gt; 1547

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6793486CD1

&lt;400&gt; 12

|   |    |    |    |
|---|----|----|----|
| Met Ser Asp Ser Leu Trp Thr Ala Leu Ser Asn Phe Ser Met Pro |    |    |    |
| 1   | 5  | 10 | 15 |
| Ser Phe Pro Gly Gly Ser Met Phe Arg Arg Thr Lys Ser Cys Arg |    |    |    |
|   | 20 | 25 | 30 |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Thr | Ser | Asn | Arg | Lys | Ser | Leu | Ile | Leu | Thr | Ser | Thr | Ser | Pro | Thr |
|     |     |     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |
| Leu | Pro | Arg | Pro | His | Ser | Pro | Leu | Pro | Gly | His | Leu | Gly | Ser | Ser |
|     |     |     |     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |
| Pro | Leu | Asp | Ser | Pro | Arg | Asn | Phe | Ser | Pro | Asn | Thr | Pro | Ala | His |
|     |     |     |     | 65  |     |     |     |     | 70  |     |     |     |     | 75  |
| Phe | Ser | Phe | Ala | Ser | Ser | Arg | Arg | Ala | Asp | Gly | Arg | Arg | Trp | Ser |
|     |     |     |     | 80  |     |     |     |     | 85  |     |     |     |     | 90  |
| Leu | Ala | Ser | Leu | Pro | Ser | Ser | Gly | Tyr | Gly | Thr | Asn | Thr | Pro | Ser |
|     |     |     |     | 95  |     |     |     |     | 100 |     |     |     |     | 105 |
| Ser | Thr | Val | Ser | Ser | Ser | Cys | Ser | Ser | Gln | Glu | Arg | Leu | His | Gln |
|     |     |     |     | 110 |     |     |     |     | 115 |     |     |     |     | 120 |
| Leu | Pro | Tyr | Gln | Pro | Thr | Val | Asp | Glu | Leu | His | Phe | Leu | Ser | Lys |
|     |     |     |     | 125 |     |     |     |     | 130 |     |     |     |     | 135 |
| His | Phe | Gly | Ser | Thr | Glu | Ser | Ile | Thr | Asp | Glu | Asp | Gly | Gly | Arg |
|     |     |     |     | 140 |     |     |     |     | 145 |     |     |     |     | 150 |
| Arg | Ser | Pro | Ala | Val | Arg | Pro | Arg | Ser | Arg | Ser | Leu | Ser | Pro | Gly |
|     |     |     |     | 155 |     |     |     |     | 160 |     |     |     |     | 165 |
| Arg | Ser | Pro | Ser | Ser | Tyr | Asp | Asn | Glu | Ile | Val | Met | Met | Asn | His |
|     |     |     |     | 170 |     |     |     |     | 175 |     |     |     |     | 180 |
| Val | Tyr | Lys | Glu | Arg | Phe | Pro | Lys | Ala | Thr | Ala | Gln | Met | Glu | Glu |
|     |     |     |     | 185 |     |     |     |     | 190 |     |     |     |     | 195 |
| Lys | Leu | Arg | Asp | Phe | Thr | Arg | Ala | Tyr | Glu | Pro | Asp | Ser | Val | Leu |
|     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |     | 210 |
| Pro | Leu | Ala | Asp | Gly | Val | Leu | Ser | Phe | Ile | His | His | Gln | Ile | Ile |
|     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     | 225 |
| Glu | Leu | Ala | Arg | Asp | Cys | Leu | Thr | Lys | Ser | Arg | Asp | Gly | Leu | Ile |
|     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
| Thr | Thr | Val | Tyr | Phe | Tyr | Glu | Leu | Gln | Glu | Asn | Leu | Glu | Lys | Leu |
|     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |
| Leu | Gln | Asp | Ala | Tyr | Glu | Arg | Ser | Glu | Ser | Leu | Glu | Val | Ala | Phe |
|     |     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |
| Val | Thr | Gln | Leu | Val | Lys | Lys | Leu | Leu | Ile | Ile | Ile | Ser | Arg | Pro |
|     |     |     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |
| Ala | Arg | Leu | Leu | Glu | Cys | Leu | Glu | Phe | Asn | Pro | Glu | Glu | Phe | Tyr |
|     |     |     |     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |
| His | Leu | Leu | Glu | Ala | Ala | Glu | Gly | His | Ala | Lys | Glu | Gly | His | Leu |
|     |     |     |     | 305 |     |     |     |     | 310 |     |     |     |     | 315 |
| Val | Lys | Thr | Asp | Ile | Pro | Arg | Tyr | Ile | Ile | Arg | Gln | Leu | Gly | Leu |
|     |     |     |     | 320 |     |     |     |     | 325 |     |     |     |     | 330 |
| Thr | Arg | Asp | Pro | Phe | Pro | Asp | Val | Val | His | Leu | Glu | Glu | Gln | Asp |
|     |     |     |     | 335 |     |     |     |     | 340 |     |     |     |     | 345 |
| Ser | Gly | Gly | Ser | Asn | Thr | Pro | Glu | Gln | Asp | Asp | Leu | Ser | Glu | Gly |
|     |     |     |     | 350 |     |     |     |     | 355 |     |     |     |     | 360 |
| Arg | Ser | Ser | Lys | Ala | Lys | Lys | Pro | Pro | Gly | Glu | Asn | Asp | Phe | Asp |
|     |     |     |     | 365 |     |     |     |     | 370 |     |     |     |     | 375 |
| Thr | Ile | Lys | Leu | Ile | Ser | Asn | Gly | Ala | Tyr | Gly | Ala | Val | Tyr | Leu |
|     |     |     |     | 380 |     |     |     |     | 385 |     |     |     |     | 390 |
| Val | Arg | His | Arg | Asp | Thr | Arg | Gln | Arg | Phe | Ala | Met | Lys | Lys | Ile |
|     |     |     |     | 395 |     |     |     |     | 400 |     |     |     |     | 405 |
| Asn | Lys | Gln | Asn | Leu | Ile | Leu | Arg | Asn | Gln | Ile | Gln | Gln | Ala | Phe |
|     |     |     |     | 410 |     |     |     |     | 415 |     |     |     |     | 420 |
| Val | Glu | Arg | Asp | Ile | Leu | Thr | Phe | Ala | Glu | Asn | Pro | Phe | Val | Val |
|     |     |     |     | 425 |     |     |     |     | 430 |     |     |     |     | 435 |
| Gly | Met | Phe | Cys | Ser | Phe | Glu | Thr | Arg | Arg | His | Leu | Cys | Met | Val |
|     |     |     |     | 440 |     |     |     |     | 445 |     |     |     |     | 450 |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Glu | Tyr | Val | Glu | Gly | Gly | Asp | Cys | Ala | Thr | Leu | Leu | Lys | Asn | 455 | 460 | 465 |
| Ile | Gly | Ala | Leu | Pro | Val | Glu | Met | Ala | Arg | Met | Tyr | Phe | Ala | Glu | 470 | 475 | 480 |
| Thr | Val | Leu | Ala | Leu | Glu | Tyr | Leu | His | Asn | Tyr | Gly | Ile | Val | His | 485 | 490 | 495 |
| Arg | Asp | Leu | Lys | Pro | Asp | Asn | Leu | Leu | Ile | Thr | Ser | Met | Gly | His | 500 | 505 | 510 |
| Ile | Lys | Leu | Thr | Asp | Phe | Gly | Leu | Ser | Lys | Met | Gly | Leu | Met | Ser | 515 | 520 | 525 |
| Leu | Thr | Thr | Asn | Leu | Tyr | Glu | Gly | His | Ile | Glu | Lys | Asp | Ala | Arg | 530 | 535 | 540 |
| Glu | Phe | Leu | Asp | Lys | Gln | Val | Cys | Gly | Thr | Pro | Glu | Tyr | Ile | Ala | 545 | 550 | 555 |
| Pro | Glu | Val | Ile | Leu | Arg | Gln | Gly | Tyr | Gly | Lys | Pro | Val | Asp | Trp | 560 | 565 | 570 |
| Trp | Ala | Met | Gly | Ile | Ile | Leu | Tyr | Glu | Phe | Leu | Val | Gly | Cys | Val | 575 | 580 | 585 |
| Pro | Phe | Phe | Gly | Asp | Thr | Pro | Glu | Glu | Leu | Phe | Gly | Gln | Val | Ile | 590 | 595 | 600 |
| Ser | Asp | Asp | Ile | Leu | Trp | Pro | Glu | Gly | Asp | Glu | Ala | Leu | Pro | Thr | 605 | 610 | 615 |
| Glu | Ala | Gln | Leu | Leu | Ile | Ser | Ser | Leu | Leu | Gln | Thr | Asn | Pro | Leu | 620 | 625 | 630 |
| Val | Arg | Leu | Gly | Ala | Gly | Gly | Ala | Phe | Glu | Val | Lys | Gln | His | Ser | 635 | 640 | 645 |
| Phe | Phe | Arg | Asp | Leu | Asp | Trp | Thr | Gly | Leu | Leu | Arg | Gln | Lys | Ala | 650 | 655 | 660 |
| Glu | Phe | Ile | Pro | His | Leu | Glu | Ser | Glu | Asp | Asp | Thr | Ser | Tyr | Phe | 665 | 670 | 675 |
| Asp | Thr | Arg | Ser | Asp | Arg | Tyr | His | His | Val | Asn | Ser | Tyr | Asp | Glu | 680 | 685 | 690 |
| Asp | Asp | Thr | Thr | Glu | Glu | Glu | Pro | Val | Glu | Ile | Arg | Gln | Phe | Ser | 695 | 700 | 705 |
| Ser | Cys | Ser | Pro | Arg | Phe | Ser | Lys | Val | Tyr | Ser | Ser | Met | Glu | Gln | 710 | 715 | 720 |
| Leu | Ser | Gln | His | Glu | Pro | Lys | Thr | Pro | Val | Ala | Ala | Ala | Gly | Ser | 725 | 730 | 735 |
| Ser | Lys | Arg | Glu | Pro | Ser | Thr | Lys | Gly | Pro | Glu | Glu | Lys | Val | Ala | 740 | 745 | 750 |
| Gly | Lys | Arg | Glu | Gly | Leu | Gly | Gly | Leu | Thr | Leu | Arg | Glu | Lys | Ser | 755 | 760 | 765 |
| Ile | Thr | Thr | Pro | Pro | Pro | Cys | Ser | Lys | Arg | Phe | Ser | Ala | Ser | Glu | 770 | 775 | 780 |
| Ala | Ser | Phe | Leu | Glu | Gly | Glu | Ala | Ser | Pro | Pro | Leu | Gly | Ala | Arg | 785 | 790 | 795 |
| Arg | Arg | Phe | Ser | Ala | Leu | Leu | Glu | Pro | Ser | Arg | Phe | Ser | Ala | Pro | 800 | 805 | 810 |
| Gln | Glu | Asp | Glu | Asp | Glu | Ala | Arg | Leu | Arg | Arg | Pro | Pro | Arg | Pro | 815 | 820 | 825 |
| Ser | Ser | Asp | Pro | Ala | Gly | Ser | Leu | Asp | Ala | Arg | Ala | Pro | Lys | Glu | 830 | 835 | 840 |
| Glu | Thr | Gln | Gly | Glu | Gly | Thr | Ser | Ser | Ala | Gly | Asp | Ser | Glu | Ala | 845 | 850 | 855 |
| Thr | Asp | Arg | Pro | Arg | Pro | Gly | Asp | Leu | Cys | Pro | Pro | Ser | Lys | Asp | 860 | 865 | 870 |

|   |      |           |
|---|------|-----------|
| Gly Asp Ala Ser Gly Pro Arg Ala Thr Asn Asp Leu Val Leu Arg |      |           |
|   | 875  | 880 885   |
| Arg Ala Arg His Gln Gln Met Ser Gly Asp Val Ala Val Glu Lys |      |           |
|   | 890  | 895 900   |
| Arg Pro Ser Arg Thr Gly Gly Lys Val Ile Lys Ser Ala Ser Ala |      |           |
|   | 905  | 910 915   |
| Thr Ala Leu Ser Val Met Ile Pro Ala Val Asp Pro His Gly Ser |      |           |
|   | 920  | 925 930   |
| Ser Pro Leu Ala Ser Pro Met Ser Pro Arg Ser Leu Ser Ser Asn |      |           |
|   | 935  | 940 945   |
| Pro Ser Ser Arg Asp Ser Ser Pro Ser Arg Asp Tyr Ser Pro Ala |      |           |
|   | 950  | 955 960   |
| Val Ser Gly Leu Arg Ser Pro Ile Thr Ile Gln Arg Ser Gly Lys |      |           |
|   | 965  | 970 975   |
| Lys Tyr Gly Phe Thr Leu Arg Ala Ile Arg Val Tyr Met Gly Asp |      |           |
|   | 980  | 985 990   |
| Thr Asp Val Tyr Ser Val His His Ile Val Trp His Val Glu Glu |      |           |
|   | 995  | 1000 1005 |
| Gly Gly Pro Ala Gln Glu Ala Gly Leu Cys Ala Gly Asp Leu Ile |      |           |
|   | 1010 | 1015 1020 |
| Thr His Val Asn Gly Glu Pro Val His Gly Met Val His Pro Glu |      |           |
|   | 1025 | 1030 1035 |
| Val Val Glu Leu Ile Leu Lys Ser Gly Asn Lys Val Ala Val Thr |      |           |
|   | 1040 | 1045 1050 |
| Thr Thr Pro Phe Glu Asn Thr Ser Ile Arg Ile Gly Pro Ala Arg |      |           |
|   | 1055 | 1060 1065 |
| Arg Ser Ser Tyr Lys Ala Lys Met Ala Arg Arg Asn Lys Arg Pro |      |           |
|   | 1070 | 1075 1080 |
| Ser Ala Lys Glu Gly Gln Glu Ser Lys Lys Arg Ser Ser Leu Phe |      |           |
|   | 1085 | 1090 1095 |
| Arg Lys Ile Thr Lys Gln Ser Asn Leu Leu His Thr Ser Arg Ser |      |           |
|   | 1100 | 1105 1110 |
| Leu Ser Ser Leu Asn Arg Ser Leu Ser Ser Ser Asp Ser Leu Pro |      |           |
|   | 1115 | 1120 1125 |
| Gly Ser Pro Thr His Gly Leu Pro Ala Arg Ser Pro Thr His Ser |      |           |
|   | 1130 | 1135 1140 |
| Tyr Arg Ser Thr Pro Asp Ser Ala Tyr Leu Gly Ile Thr Ser Cys |      |           |
|   | 1145 | 1150 1155 |
| Thr Cys Ala Gly Thr Glu Gln Thr Pro Asn Ser Pro Ala Ser Ser |      |           |
|   | 1160 | 1165 1170 |
| Ala Ser His His Ile Arg Pro Ser Thr Leu His Gly Leu Ser Pro |      |           |
|   | 1175 | 1180 1185 |
| Lys Leu His Arg Gln Tyr Arg Ser Ala Arg Cys Lys Ser Ala Gly |      |           |
|   | 1190 | 1195 1200 |
| Asn Ile Pro Leu Ser Pro Leu Ala His Thr Pro Ser Pro Thr Gln |      |           |
|   | 1205 | 1210 1215 |
| Ala Ser Pro Pro Pro Leu Pro Gly His Thr Arg Pro Lys Ser Ala |      |           |
|   | 1220 | 1225 1230 |
| Glu Pro Pro Arg Ser Pro Leu Leu Lys Arg Val Gln Ser Ala Glu |      |           |
|   | 1235 | 1240 1245 |
| Lys Leu Gly Ala Ser Leu Ser Ala Asp Lys Lys Gly Ala Leu Arg |      |           |
|   | 1250 | 1255 1260 |
| Lys His Ser Leu Glu Val Gly His Pro Asp Phe Arg Lys Asp Phe |      |           |
|   | 1265 | 1270 1275 |
| His Gly Glu Leu Ala Leu His Ser Leu Ala Glu Ser Asp Gly Glu |      |           |
|   | 1280 | 1285 1290 |

Thr Pro Pro Val Glu Gly Leu Gly Ala Pro Arg Gln Val Ala Val  
 1295 1300 1305  
 Arg Arg Leu Gly Arg Gln Glu Ser Pro Leu Ser Leu Gly Ala Asp  
 1310 1315 1320  
 Pro Leu Leu Pro Glu Gly Ala Ser Arg Pro Pro Val Ser Ser Lys  
 1325 1330 1335  
 Glu Lys Glu Ser Pro Gly Gly Ala Glu Ala Cys Thr Pro Pro Arg  
 1340 1345 1350  
 Ala Thr Thr Pro Gly Gly Arg Thr Leu Glu Arg Asp Val Gly Cys  
 1355 1360 1365  
 Thr Arg His Gln Ser Val Gln Thr Glu Asp Gly Thr Gly Gly Met  
 1370 1375 1380  
 Ala Arg Ala Val Ala Lys Ala Ala Leu Ser Pro Val Gln Glu His  
 1385 1390 1395  
 Glu Thr Gly Arg Arg Ser Ser Ser Gly Glu Ala Gly Thr Pro Leu  
 1400 1405 1410  
 Val Pro Ile Val Val Glu Pro Ala Arg Pro Gly Ala Lys Ala Val  
 1415 1420 1425  
 Val Pro Gln Pro Leu Gly Ala Asp Ser Lys Gly Leu Gln Glu Pro  
 1430 1435 1440  
 Ala Pro Leu Ala Pro Ser Val Pro Glu Ala Pro Arg Gly Arg Glu  
 1445 1450 1455  
 Arg Trp Val Leu Glu Val Val Glu Glu Arg Thr Thr Leu Ser Gly  
 1460 1465 1470  
 Pro Arg Ser Lys Pro Ala Ser Pro Lys Leu Ser Pro Glu Pro Gln  
 1475 1480 1485  
 Thr Pro Ser Leu Ala Pro Ala Lys Cys Ser Ala Pro Ser Ser Ala  
 1490 1495 1500  
 Val Thr Pro Val Pro Pro Ala Ser Leu Leu Gly Ser Gly Thr Lys  
 1505 1510 1515  
 Pro Gln Val Gly Leu Thr Ser Arg Cys Pro Ala Glu Ala Val Pro  
 1520 1525 1530  
 Pro Ala Gly Leu Thr Lys Lys Gly Val Ser Ser Pro Ala Pro Pro  
 1535 1540 1545  
 Gly Pro

&lt;210&gt; 13

&lt;211&gt; 505

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7494178CD1

&lt;400&gt; 13

Met Leu Asn Arg Val Arg Ser Ala Val Ala His Leu Val Ser Ser  
 1 5 10 15  
 Gly Gly Ala Pro Pro Pro Arg Pro Lys Ser Pro Asp Leu Pro Asn  
 20 25 30  
 Ala Ala Ser Ala Pro Pro Ala Ala Ala Pro Glu Ala Pro Arg Ser  
 35 40 45  
 Pro Pro Ala Lys Ala Gly Ser Gly Ser Ala Thr Pro Ala Lys Ala  
 50 55 60  
 Val Glu Ala Arg Ala Ser Phe Ser Arg Pro Thr Phe Leu Gln Leu



|   |     |     |
|---|-----|-----|
| 65  | 70  | 75  |
| Ser Pro Gly Gly Leu Arg Arg Ala Asp Asp His Ala Gly Arg Ala |     |     |
| 80  | 85  | 90  |
| Val Gln Ser Pro Pro Asp Thr Gly Arg Arg Leu Pro Trp Ser Thr |     |     |
| 95  | 100 | 105 |
| Gly Tyr Ala Glu Val Ile Asn Ala Gly Lys Ser Arg His Asn Glu |     |     |
| 110   | 115 | 120 |
| Asp Gln Ala Cys Cys Glu Val Val Tyr Val Glu Gly Arg Arg Ser |     |     |
| 125   | 130 | 135 |
| Val Thr Gly Val Pro Arg Glu Pro Ser Arg Gly Gln Gly Leu Cys |     |     |
| 140   | 145 | 150 |
| Phe Tyr Tyr Trp Gly Leu Phe Asp Gly His Ala Gly Gly Gly Ala |     |     |
| 155   | 160 | 165 |
| Ala Glu Met Ala Ser Arg Leu Leu His Arg His Ile Arg Glu Gln |     |     |
| 170   | 175 | 180 |
| Leu Lys Asp Leu Val Glu Ile Leu Gln Asp Pro Ser Pro Pro Pro |     |     |
| 185   | 190 | 195 |
| Leu Cys Leu Pro Thr Thr Pro Gly Thr Pro Asp Ser Ser Asp Pro |     |     |
| 200   | 205 | 210 |
| Ser His Leu Leu Gly Pro Gln Ser Cys Trp Ser Ser Gln Lys Glu |     |     |
| 215   | 220 | 225 |
| Val Ser His Glu Ser Leu Val Val Gly Ala Ile Glu Asn Ala Phe |     |     |
| 230   | 235 | 240 |
| Gln Leu Met Asp Glu Gln Met Ala Arg Glu Arg Arg Gly His Gln |     |     |
| 245   | 250 | 255 |
| Val Glu Gly Gly Cys Cys Ala Leu Val Val Ile Tyr Leu Leu Gly |     |     |
| 260   | 265 | 270 |
| Lys Val Tyr Val Ala Asn Ala Gly Asp Ser Arg Ala Ile Ile Val |     |     |
| 275   | 280 | 285 |
| Arg Asn Gly Glu Ile Ile Pro Met Ser Arg Glu Phe Thr Pro Glu |     |     |
| 290   | 295 | 300 |
| Thr Glu Arg Gln Arg Leu Gln Leu Leu Gly Phe Leu Lys Pro Glu |     |     |
| 305   | 310 | 315 |
| Leu Leu Gly Ser Glu Phe Thr His Leu Glu Phe Pro Arg Arg Val |     |     |
| 320   | 325 | 330 |
| Leu Pro Lys Glu Leu Gly Gln Arg Met Leu Tyr Arg Asp Gln Asn |     |     |
| 335   | 340 | 345 |
| Met Thr Gly Trp Ala Tyr Lys Lys Ile Glu Leu Glu Asp Leu Arg |     |     |
| 350   | 355 | 360 |
| Phe Pro Leu Val Cys Gly Glu Gly Lys Lys Ala Arg Val Met Ala |     |     |
| 365   | 370 | 375 |
| Thr Ile Gly Val Thr Arg Gly Leu Gly Asp His Ser Leu Lys Val |     |     |
| 380   | 385 | 390 |
| Cys Ser Ser Thr Leu Pro Ile Lys Pro Phe Leu Ser Cys Phe Pro |     |     |
| 395   | 400 | 405 |
| Glu Val Arg Val Tyr Asp Leu Thr Gln Tyr Glu His Cys Pro Asp |     |     |
| 410   | 415 | 420 |
| Asp Val Leu Val Leu Gly Thr Asp Gly Leu Trp Asp Val Thr Thr |     |     |
| 425   | 430 | 435 |
| Asp Cys Glu Val Ala Ala Thr Val Asp Arg Val Leu Ser Ala Tyr |     |     |
| 440   | 445 | 450 |
| Glu Pro Asn Asp His Ser Arg Tyr Thr Ala Leu Ala Gln Ala Leu |     |     |
| 455   | 460 | 465 |
| Val Leu Gly Ala Arg Gly Thr Pro Arg Asp Arg Gly Trp Arg Leu |     |     |
| 470   | 475 | 480 |
| Pro Asn Asn Lys Leu Gly Ser Gly Asp Asp Ile Ser Val Phe Val |     |     |

```
<210> 14
<211> 1036
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> misc_feature
<223> Incyte ID No: 7096516CD1
```

| <400> 14 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
| Met      | Gly | Gly | Cys | Glu | Val | Arg | Glu | Phe | Leu | Leu | Gln | Phe | Gly | Phe |  |
| 1        |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |  |
| Phe      | Leu | Pro | Leu | Leu | Thr | Ala | Trp | Pro | Gly | Asp | Cys | Ser | His | Val |  |
|          |     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |  |
| Ser      | Asn | Asn | Gln | Val | Val | Leu | Leu | Asp | Thr | Thr | Thr | Val | Leu | Gly |  |
|          |     |     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |  |
| Glu      | Leu | Gly | Trp | Lys | Thr | Tyr | Pro | Leu | Asn | Gly | Trp | Asp | Ala | Ile |  |
|          |     |     |     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |  |
| Thr      | Glu | Met | Asp | Glu | His | Asn | Arg | Pro | Ile | His | Thr | Tyr | Gln | Val |  |
|          |     |     |     | 65  |     |     |     |     | 70  |     |     |     |     | 75  |  |
| Cys      | Asn | Val | Met | Glu | Pro | Asn | Gln | Asn | Asn | Trp | Leu | Arg | Thr | Asn |  |
|          |     |     |     | 80  |     |     |     |     | 85  |     |     |     |     | 90  |  |
| Trp      | Ile | Ser | Arg | Asp | Ala | Ala | Gln | Lys | Ile | Tyr | Val | Glu | Met | Lys |  |
|          |     |     |     | 95  |     |     |     |     | 100 |     |     |     |     | 105 |  |
| Phe      | Thr | Leu | Arg | Asp | Cys | Asn | Ser | Ile | Pro | Trp | Val | Leu | Gly | Thr |  |
|          |     |     |     | 110 |     |     |     |     | 115 |     |     |     |     | 120 |  |
| Cys      | Lys | Glu | Thr | Phe | Asn | Leu | Phe | Tyr | Met | Glu | Ser | Asp | Glu | Ser |  |
|          |     |     |     | 125 |     |     |     |     | 130 |     |     |     |     | 135 |  |
| His      | Gly | Ile | Lys | Phe | Lys | Pro | Asn | Gln | Tyr | Thr | Lys | Ile | Asp | Thr |  |
|          |     |     |     | 140 |     |     |     |     | 145 |     |     |     |     | 150 |  |
| Ile      | Ala | Ala | Asp | Glu | Ser | Phe | Thr | Gln | Met | Asp | Leu | Gly | Asp | Arg |  |
|          |     |     |     | 155 |     |     |     |     | 160 |     |     |     |     | 165 |  |
| Ile      | Leu | Lys | Leu | Asn | Thr | Glu | Ile | Arg | Glu | Val | Gly | Pro | Ile | Glu |  |
|          |     |     |     | 170 |     |     |     |     | 175 |     |     |     |     | 180 |  |
| Arg      | Lys | Gly | Phe | Tyr | Leu | Ala | Phe | Gln | Asp | Ile | Gly | Ala | Cys | Ile |  |
|          |     |     |     | 185 |     |     |     |     | 190 |     |     |     |     | 195 |  |
| Ala      | Leu | Val | Ser | Val | Arg | Val | Phe | Tyr | Lys | Lys | Cys | Pro | Phe | Thr |  |
|          |     |     |     | 200 |     |     |     |     | 205 |     |     |     |     | 210 |  |
| Val      | Arg | Asn | Leu | Ala | Met | Phe | Pro | Asp | Thr | Ile | Pro | Arg | Val | Asp |  |
|          |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     | 225 |  |
| Ser      | Ser | Ser | Leu | Val | Glu | Val | Arg | Gly | Ser | Cys | Val | Lys | Ser | Ala |  |
|          |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |  |
| Glu      | Glu | Arg | Asp | Thr | Pro | Lys | Leu | Tyr | Cys | Gly | Ala | Asp | Gly | Asp |  |
|          |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |  |
| Trp      | Leu | Val | Pro | Leu | Gly | Arg | Cys | Ile | Cys | Ser | Thr | Gly | Tyr | Glu |  |
|          |     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |  |
| Glu      | Ile | Glu | Gly | Ser | Cys | His | Ala | Cys | Arg | Pro | Gly | Phe | Tyr | Lys |  |
|          |     |     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |  |
| Ala      | Phe | Ala | Gly | Asn | Thr | Lys | Cys | Ser | Lys | Cys | Pro | Pro | His | Ser |  |
|          |     |     |     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |  |
| Leu      | Thr | Tyr | Met | Glu | Ala | Thr | Ser | Val | Cys | Gln | Cys | Glu | Lys | Gly |  |
|          |     |     |     | 305 |     |     |     |     | 310 |     |     |     |     | 315 |  |

|                 |                     |                         |     |
|-----------------|---------------------|-------------------------|-----|
| Tyr Phe Arg Ala | Glu Lys Asp Pro Pro | Ser Met Ala Cys Thr Arg |     |
|                 | 320                 | 325                     | 330 |
| Pro Pro Ser Ala | Pro Arg Asn Val Val | Phe Asn Ile Asn Glu Thr |     |
|                 | 335                 | 340                     | 345 |
| Ala Leu Ile Leu | Glu Trp Ser Pro Pro | Ser Asp Thr Gly Gly Arg |     |
|                 | 350                 | 355                     | 360 |
| Lys Asp Leu Thr | Tyr Ser Val Ile Cys | Lys Lys Cys Gly Leu Asp |     |
|                 | 365                 | 370                     | 375 |
| Thr Ser Gln Cys | Glu Asp Cys Gly Gly | Gly Leu Arg Phe Ile Pro |     |
|                 | 380                 | 385                     | 390 |
| Arg His Thr Gly | Leu Ile Asn Asn Ser | Val Ile Val Leu Asp Phe |     |
|                 | 395                 | 400                     | 405 |
| Val Ser His Val | Asn Tyr Thr Phe Glu | Ile Glu Ala Met Asn Gly |     |
|                 | 410                 | 415                     | 420 |
| Val Ser Glu Leu | Ser Phe Ser Pro Lys | Pro Phe Thr Ala Ile Thr |     |
|                 | 425                 | 430                     | 435 |
| Val Thr Thr Asp | Gln Asp Ala Pro Ser | Leu Ile Gly Val Val Arg |     |
|                 | 440                 | 445                     | 450 |
| Lys Asp Trp Ala | Ser Gln Asn Ser Ile | Ala Leu Ser Trp Gln Ala |     |
|                 | 455                 | 460                     | 465 |
| Pro Ala Phe Ser | Asn Gly Ala Ile Leu | Asp Tyr Glu Ile Lys Tyr |     |
|                 | 470                 | 475                     | 480 |
| Tyr Glu Lys Glu | His Glu Gln Leu Thr | Tyr Ser Ser Thr Arg Ser |     |
|                 | 485                 | 490                     | 495 |
| Lys Ala Pro Ser | Val Ile Ile Thr Gly | Leu Lys Pro Ala Thr Lys |     |
|                 | 500                 | 505                     | 510 |
| Tyr Val Phe His | Ile Arg Val Arg Thr | Ala Thr Gly Tyr Ser Gly |     |
|                 | 515                 | 520                     | 525 |
| Tyr Ser Gln Lys | Phe Glu Phe Glu Thr | Gly Asp Glu Thr Ser Asp |     |
|                 | 530                 | 535                     | 540 |
| Met Ala Ala Glu | Gln Gly Gln Ile Leu | Val Ile Ala Thr Ala Ala |     |
|                 | 545                 | 550                     | 555 |
| Val Gly Gly Phe | Thr Leu Leu Val Ile | Leu Thr Leu Phe Phe Leu |     |
|                 | 560                 | 565                     | 570 |
| Ile Thr Gly Arg | Cys Gln Trp Tyr Ile | Lys Ala Lys Met Lys Ser |     |
|                 | 575                 | 580                     | 585 |
| Glu Glu Lys Arg | Arg Asn His Leu Gln | Asn Gly His Leu Arg Phe |     |
|                 | 590                 | 595                     | 600 |
| Pro Gly Ile Lys | Thr Tyr Ile Asp Pro | Asp Thr Tyr Glu Asp Pro |     |
|                 | 605                 | 610                     | 615 |
| Ser Leu Ala Val | His Glu Phe Ala Lys | Glu Ile Asp Pro Ser Arg |     |
|                 | 620                 | 625                     | 630 |
| Ile Arg Ile Glu | Arg Val Ile Gly Ala | Gly Glu Phe Gly Glu Val |     |
|                 | 635                 | 640                     | 645 |
| Cys Ser Gly Arg | Leu Lys Thr Pro Gly | Lys Arg Glu Ile Pro Val |     |
|                 | 650                 | 655                     | 660 |
| Ala Ile Lys Thr | Leu Lys Gly Gly His | Met Asp Arg Gln Arg Arg |     |
|                 | 665                 | 670                     | 675 |
| Asp Phe Leu Arg | Glu Ala Ser Ile Met | Gly Gln Phe Asp His Pro |     |
|                 | 680                 | 685                     | 690 |
| Asn Ile Ile Arg | Leu Glu Gly Val Val | Thr Lys Arg Ser Phe Pro |     |
|                 | 695                 | 700                     | 705 |
| Ala Ile Gly Val | Glu Ala Phe Cys Pro | Ser Phe Leu Arg Ala Gly |     |
|                 | 710                 | 715                     | 720 |
| Phe Leu Asn Ser | Ile Gln Ala Pro His | Pro Val Pro Gly Gly Gly |     |
|                 | 725                 | 730                     | 735 |

```

Ser Leu Pro Pro Arg Ile Pro Ala Gly Arg Pro Val Met Ile Val
      740                      745                      750
Val Glu Tyr Met Glu Asn Gly Ser Leu Asp Ser Phe Leu Arg Lys
      755                      760                      765
His Asp Gly His Phe Thr Val Ile Gln Leu Val Gly Met Leu Arg
      770                      775                      780
Gly Ile Ala Ser Gly Met Lys Tyr Leu Ser Asp Met Gly Tyr Val
      785                      790                      795
His Arg Asp Leu Ala Ala Arg Asn Ile Leu Val Asn Ser Asn Leu
      800                      805                      810
Val Cys Lys Val Ser Asp Phe Gly Leu Ser Arg Val Leu Glu Asp
      815                      820                      825
Asp Pro Glu Ala Ala Tyr Thr Thr Thr Gly Gly Lys Ile Pro Ile
      830                      835                      840
Arg Trp Thr Ala Pro Glu Ala Ile Ala Tyr Arg Lys Phe Ser Ser
      845                      850                      855
Ala Ser Asp Ala Trp Ser Tyr Gly Ile Val Met Trp Glu Val Met
      860                      865                      870
Ser Tyr Gly Glu Arg Pro Tyr Trp Glu Met Ser Asn Gln Asp Val
      875                      880                      885
Ile Leu Ser Ile Glu Glu Gly Tyr Arg Leu Pro Ala Pro Met Gly
      890                      895                      900
Cys Pro Ala Ser Leu His Gln Leu Met Leu His Cys Trp Gln Lys
      905                      910                      915
Glu Arg Asn His Arg Pro Lys Phe Thr Asp Ile Val Ser Phe Leu
      920                      925                      930
Asp Lys Leu Ile Arg Asn Pro Ser Ala Leu His Thr Leu Val Glu
      935                      940                      945
Asp Ile Leu Val Met Pro Glu Ser Pro Gly Glu Val Pro Glu Tyr
      950                      955                      960
Pro Leu Phe Val Thr Val Gly Asp Trp Leu Asp Ser Ile Lys Met
      965                      970                      975
Gly Gln Tyr Lys Asn Asn Phe Val Ala Ala Gly Phe Thr Thr Phe
      980                      985                      990
Asp Leu Ile Ser Arg Met Ser Ile Asp Asp Ile Arg Arg Ile Gly
      995                      1000                     1005
Val Ile Leu Ile Gly His Gln Arg Arg Ile Val Ser Ser Ile Gln
      1010                     1015                     1020
Thr Leu Arg Leu His Met Met His Ile Gln Glu Glu Gly Phe His
      1025                     1030                     1035
Val

```

&lt;210&gt; 15

&lt;211&gt; 489

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7474666CD1

&lt;400&gt; 15

```

Met Glu Pro Gly Leu Glu His Ala Leu Arg Arg Thr Pro Ser Trp
  1           5           10           15
Ser Ser Leu Gly Gly Ser Glu His Gln Glu Met Ser Phe Leu Glu

```

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|     | 20  |     | 25  |     | 30  |     |     |     |     |     |     |     |     |     |
| Gln | Glu | Asn | Ser | Ser | Ser | Trp | Pro | Ser | Pro | Ala | Val | Thr | Ser | Ser |
|     | 35  |     |     |     |     |     |     |     | 40  |     |     |     |     | 45  |
| Ser | Glu | Arg | Ile | Arg | Gly | Lys | Arg | Arg | Ala | Lys | Ala | Leu | Arg | Trp |
|     | 50  |     |     |     |     |     |     |     | 55  |     |     |     |     | 60  |
| Thr | Arg | Gln | Lys | Ser | Val | Glu | Glu | Gly | Glu | Pro | Pro | Gly | Gln | Gly |
|     | 65  |     |     |     |     |     |     |     | 70  |     |     |     |     | 75  |
| Glu | Gly | Pro | Arg | Ser | Arg | Pro | Ala | Ala | Glu | Ser | Thr | Gly | Leu | Glu |
|     | 80  |     |     |     |     |     |     |     | 85  |     |     |     |     | 90  |
| Ala | Thr | Phe | Pro | Lys | Thr | Thr | Pro | Leu | Ala | Gln | Ala | Asp | Pro | Ala |
|     | 95  |     |     |     |     |     |     |     | 100 |     |     |     |     | 105 |
| Gly | Val | Gly | Thr | Pro | Pro | Thr | Gly | Trp | Asp | Cys | Leu | Pro | Ser | Asp |
|     | 110 |     |     |     |     |     |     |     | 115 |     |     |     |     | 120 |
| Cys | Thr | Ala | Ser | Ala | Ala | Gly | Ser | Ser | Thr | Asp | Asp | Val | Glu | Leu |
|     | 125 |     |     |     |     |     |     |     | 130 |     |     |     |     | 135 |
| Ala | Thr | Glu | Phe | Pro | Ala | Thr | Glu | Ala | Trp | Glu | Cys | Glu | Leu | Glu |
|     | 140 |     |     |     |     |     |     |     | 145 |     |     |     |     | 150 |
| Gly | Leu | Leu | Glu | Glu | Arg | Pro | Ala | Leu | Cys | Leu | Ser | Pro | Gln | Ala |
|     | 155 |     |     |     |     |     |     |     | 160 |     |     |     |     | 165 |
| Pro | Phe | Pro | Lys | Leu | Gly | Trp | Asp | Asp | Glu | Leu | Arg | Lys | Pro | Gly |
|     | 170 |     |     |     |     |     |     |     | 175 |     |     |     |     | 180 |
| Ala | Gln | Ile | Tyr | Met | Arg | Phe | Met | Gln | Glu | His | Thr | Cys | Tyr | Asp |
|     | 185 |     |     |     |     |     |     |     | 190 |     |     |     |     | 195 |
| Ala | Met | Ala | Thr | Ser | Ser | Lys | Leu | Val | Ile | Phe | Asp | Thr | Met | Leu |
|     | 200 |     |     |     |     |     |     |     | 205 |     |     |     |     | 210 |
| Glu | Ile | Lys | Lys | Ala | Phe | Phe | Ala | Leu | Val | Ala | Asn | Gly | Val | Arg |
|     | 215 |     |     |     |     |     |     |     | 220 |     |     |     |     | 225 |
| Ala | Ala | Pro | Leu | Trp | Asp | Ser | Lys | Lys | Gln | Ser | Phe | Val | Gly | Met |
|     | 230 |     |     |     |     |     |     |     | 235 |     |     |     |     | 240 |
| Leu | Thr | Ile | Thr | Asp | Phe | Ile | Leu | Val | Leu | His | Arg | Tyr | Tyr | Arg |
|     | 245 |     |     |     |     |     |     |     | 250 |     |     |     |     | 255 |
| Ser | Pro | Leu | Val | Gln | Ile | Tyr | Glu | Ile | Glu | Gln | His | Lys | Ile | Glu |
|     | 260 |     |     |     |     |     |     |     | 265 |     |     |     |     | 270 |
| Thr | Trp | Arg | Glu | Ile | Tyr | Leu | Gln | Gly | Cys | Phe | Lys | Pro | Leu | Val |
|     | 275 |     |     |     |     |     |     |     | 280 |     |     |     |     | 285 |
| Ser | Ile | Ser | Pro | Asn | Asp | Ser | Leu | Phe | Glu | Ala | Val | Tyr | Thr | Leu |
|     | 290 |     |     |     |     |     |     |     | 295 |     |     |     |     | 300 |
| Ile | Lys | Asn | Arg | Ile | His | Arg | Leu | Pro | Val | Leu | Asp | Pro | Val | Ser |
|     | 305 |     |     |     |     |     |     |     | 310 |     |     |     |     | 315 |
| Gly | Asn | Val | Leu | His | Ile | Leu | Thr | His | Lys | Arg | Leu | Leu | Lys | Phe |
|     | 320 |     |     |     |     |     |     |     | 325 |     |     |     |     | 330 |
| Leu | His | Ile | Phe | Gly | Ser | Leu | Leu | Pro | Arg | Pro | Ser | Phe | Leu | Tyr |
|     | 335 |     |     |     |     |     |     |     | 340 |     |     |     |     | 345 |
| Arg | Thr | Ile | Gln | Asp | Leu | Gly | Ile | Gly | Thr | Phe | Arg | Asp | Leu | Ala |
|     | 350 |     |     |     |     |     |     |     | 355 |     |     |     |     | 360 |
| Val | Val | Leu | Glu | Thr | Ala | Pro | Ile | Leu | Thr | Ala | Leu | Asp | Ile | Phe |
|     | 365 |     |     |     |     |     |     |     | 370 |     |     |     |     | 375 |
| Val | Asp | Arg | Arg | Val | Ser | Ala | Leu | Pro | Val | Val | Asn | Glu | Cys | Gly |
|     | 380 |     |     |     |     |     |     |     | 385 |     |     |     |     | 390 |
| Gln | Val | Val | Gly | Leu | Tyr | Ser | Arg | Phe | Asp | Val | Ile | His | Leu | Ala |
|     | 395 |     |     |     |     |     |     |     | 400 |     |     |     |     | 405 |
| Ala | Gln | Gln | Thr | Tyr | Asn | His | Leu | Asp | Met | Ser | Val | Gly | Glu | Ala |
|     | 410 |     |     |     |     |     |     |     | 415 |     |     |     |     | 420 |
| Leu | Arg | Gln | Arg | Thr | Leu | Cys | Leu | Glu | Gly | Val | Leu | Ser | Cys | Gln |
|     | 425 |     |     |     |     |     |     |     | 430 |     |     |     |     | 435 |
| Pro | His | Glu | Ser | Leu | Gly | Glu | Val | Ile | Asp | Arg | Ile | Ala | Arg | Glu |

|   |     |  |     |  |     |
|---|-----|--|-----|--|-----|
|   | 440 |  | 445 |  | 450 |
| Gln Val His Arg Leu Val Leu Val Asp Glu Thr Gln His Leu Leu |     |  |     |  |     |
|   | 455 |  | 460 |  | 465 |
| Gly Val Val Ser Leu Ser Asp Ile Leu Gln Ala Leu Val Leu Ser |     |  |     |  |     |
|   | 470 |  | 475 |  | 480 |
| Pro Ala Gly Ile Asp Ala Leu Gly Ala                         |     |  |     |  |     |
|   | 485 |  |     |  |     |

&lt;210&gt; 16

&lt;211&gt; 4188

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2763993CB1

&lt;400&gt; 16

```

atgaagaagt tctctcggat gccaagtcg gagggcgcca gcggcgccgg agcgcgccgg 60
ggcgccggctg gcggggccgg ggccggggcc ggctgcggct ccggcgccgc gtccgtgggg 120
gtccgggtgt tcgcggtcgg ccgccaccag gtcaccctgg aagagtcgct ggccgaaggt 180
ggattctcca cagttttctt cgtgcgtact cagggtggaa tccgatgtgc attgaagcga 240
atgtatgtca ataacatgcc agacctcaat gtttgtaaaa gggaaattac aattatgaaa 300
gagctatctg gtcacaaaaa tattgtgggc tatttggact gtgctgttaa ttcaattagt 360
gataatgtat gggaagtctt tatcttaatg gaattattgtc gagctggaca ggtagtgaat 420
caaatgaata agaagctaca gacgggtttt acagaaccag aagtgttaca gatattctgt 480
gatacctgtg aagctgttgc aaggttgcac cagtgttaaga ctccaataat tcaccgggat 540
ctgaaggtag aaaatatctt gttgaatgat ggtgggaact atgtactttg tgactttggc 600
agtgcacta ataaatttct taatcctcaa aaagatggag ttaatgtagt agaagaagaa 660
attaaaaagt atacaactct gtcatacaga gccctgaaa tgatcaacct ttatggaggg 720
aaacctatca ccaccaaggc tgatatctgg gcaactggat gtctactcta taaactttgt 780
ttcttactc ttcttttgg tgagagtcag gttgctatct gtgatggcaa cttcaccatc 840
ccagacaatt ctctgttact ccgtaacata cattgcttaa taaggttcat gcttgaacca 900
gatccggaac atagacctga tatatttcaa gtgtcatatt ttgcatttaa atttgccaaa 960
aaggattgtc cagtctcaa catcaataat tcttctattc cttcagctct tcctgaaccg 1020
atgactgtca gtgaagcagc tgctaggaaa agccaaataa aagccagaat aacagatacc 1080
attggacca cagaaacctc aattgcacca agacaaagac caaaggccaa ctctgtctact 1140
actgccactc ccagtgtgct gaccattcaa agttcagcaa cacctgttaa agtccttgct 1200
cctggtgaat tcggtaacca tagaccaaa ggggcactaa gacctggaaa tggccctgaa 1260
attttattgg gtcagggacc tcctcagcag ccgccacagc agcatagagt actccagcaa 1320
ctacagcagg gagattggag attacagcaa ctccatttac agcatcgtca tcctcaccag 1380
cagcagcagc agcagcagca gcaacagcaa cagcagcagc agcaacagca acagcagcag 1440
cagcagcagc agcagcagca ccaccaccac caccaccacc acctacttca agatgcttat 1500
atgcagcagt atcaacatgc aacacagcag caacagatgc ttcaacaaca atttttaatg 1560
cattcgggtat atcaaccaca accttctgca tcacagtatc ctacaatgat gccgcagtat 1620
cagcaggctt tctttcaaca gcagatgcta gctcaacatc agccgtctca acaacaggca 1680
tcacctgaat atcttacctc ccctcaagag ttctcaccag ccttagtttc ctacacttca 1740
tcaacttcag ctcagggttg aaccataatg gactcctcct atagtgccaa taggtcagtt 1800
gctgataaag aggccattgc aaatttcaca aatcagaaga acatcagcaa tccacctgat 1860
atgtcagggt ggaatccttt tggagaggat aatttctcta agttaacaga agaggaaacta 1920
ttggacagag aatttgacct tctaagatca aataggtcgc aggagagagc atcctcagat 1980
aagaatgtag actcactttc tgetccacat aaccatcctc cagaagatcc ttttggttct 2040
gttcttttca tttctcattc aggttctcct gaaaagaaag ctgaacattc atctataaat 2100
caagaaaatg gcaactgcaa ccctatcaag aacggtaaaa caagtccagc atctaaagat 2160
cagcggactg gaaagaaaac ctcagtacag ggtcaagtgc aaaaggggaa tgatgaatct 2220
gaaagtgatt ttgaatcaga tcccccttct cctaagagca gtgaagagga agagcaagat 2280

```

```

gatgaagaag ttcttcaggg ggaacaagga gattttaatg atgatgatac tgaaccagaa 2340
aatctgggtc ataggcctct cctcatggat tctgaagatg aggaagaaga ggagaaacat 2400
agctctgatt ctgattatga gcaggctaaa gcaaggtaca gtgacatgag ctctgtctac 2460
agagacagat ctggcagtg gacaaaccaa gatcttaata caatactcct cacctcagcc 2520
caattatcct ctgatgttgc agtggagact cccaaacagg agtttgatgt atttggcgct 2580
gtccccctct ttgcagtgcg tgcctcaacag cccagcaag aaaagaatga aaagaacctc 2640
cctcaacaca ggtttcctgc tgcaggactg gagcaggagg aatttgatgt attcacaag 2700
gcgcctttta gcaagaaggt gaatgtacaa gaatgccatg cagtggggcc tgaggcacat 2760
actatccctg gttatcccaa aagtgtagat gtatttggct ccactccatt tcagcccttc 2820
ctcacatcaa caagtaaaag tgaaagcaat gaggaccttt ttgggcttgt gccctttgat 2880
gaaataacgg ggagccagca gcaaaaagtc aaacagcgca cgttacagaa actgtcctct 2940
cgccaaaggc gcacaaagca ggatatgtcc aaaagtaatg ggaagcggca tcatggcagc 3000
ccaactagca caaagaagac tttgaagcct acctatcgca ctccagagag ggctcgcagg 3060
cacaataaag tgggcccgcg agactctcaa agtagcaatg aatttttaac catctcagac 3120
tccaaggaga acattagtgt tgcactgact gatgggaaag ataggggaa atgtcttaca 3180
cctgaggaga gcctgttggg ccccttcggt gccagccct tccattctcc agacctgtca 3240
tggcaccttc cacatcaggg cctgagcgac atccgtgctg atcacaatac tgtcctgcca 3300
gggcgcccaa gacaaaattc actacatggg tcattccata gtgcagatgt attgaaaatg 3360
gatgatattg gtgccgtgcc ctttacagaa cttgtggtgc aaagcatcac tccacatcag 3420
tcccaacagt cccaaccagt cgaattagac ccatttgggtg ctgtccatt tccttctaaa 3480
cagtagatac ttctgatgga ttctcgcat taactcctgt ttcaaaaaag tgtgaacagt 3540
tttatgaatt tgaaagaaaa tttggtagct ctttatagca ttattctta aagatcagtc 3600
agaataggtg atttctaaat aaaccaata gaagaatgaa gtatctctac agggtagtaa 3660
cttgattcct cttcaggaga aaaggagct aaattgcaag ctctaactaa ggggttctgc 3720
tactgacatc acaacacaga aatgcaagt tggtacttcc agtgaaagca catggcacct 3780
ttctaggtgt gtagccactg agaagggaca gtgaaactgt tatttttgat atcagaatgt 3840
catttttatg tgcataatccc taaaattagg gttatttcta catacactag ttacacttgt 3900
gaattttttt taaggctctt ttttaatttc agacagttaa aaacaatcta gttatcttaa 3960
agcattagaa agttattatc tggagagtgc agagatttca gtccatacac ctttctccac 4020
aagcagagcc agaagtaact gactattgtg cctaaaactc tgtttcattt ttaaaaacaa 4080
gtgccattaa aatggaatat ctaatgataa gcatatgaaa taatgtgtaa ttagctcaat 4140
ttaactattc cacaacttac atattccaaa caatgttata catgataa 4188

```

<210> 17  
 <211> 4675  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 3684162CB1

```

<400> 17
agacaaaggg cggtcgcgc cggggccgc acgtctcgg gctctgcctc gggaaggaga 60
cttggctctga aagatgccac attcctgcag cctctcttgg tgcagtggaa tacagtcttg 120
ggcaggtgg cgtggatgag ctggtgaaag aggatgctgc ccacatccaa aggtccaga 180
ggatcctggc ctggcagctg agctcccctg catttggaaac ctcaggcgta acttgggtgta 240
gagctcatga aagtgcttg tgtttctcca gcttttttca ccagtgttca ccagactggc 300
tcaggttttg gaatctaagg tgagctggta gaaacaggag aggtagaaag aagcccctgg 360
atgcctccag aattcattga tgggatccct gcatactgct tggaaacacag aaagaggctg 420
tgacacagct gagctttgga gcatcttaag gagctcagct cagcaaacaa ctcttgcat 480
tcagccagaa agagcctctt gtaacaaagt attcaaaggg gagagtttct gcatctttta 540
ctttgcagtc cactatggta gaaaacttga cattccatag ataattgatac tgggttttct 600
ttccaagatg ccagctttta aagaaatag agccattcta agctttaaga agggttcagg 660
aaacacagga attagtagac agccctccca atgcagggtta agacgacagc ctgcgcccc 720
aactagcaca gctcagcgag catgaccata tgccattctc gtctccagag agctgggtggc 780

```

```

agtgacctca ctaggagaaa acacatccct cagccgtggg acttgacaga atgaggtgcg 840
cgagggaggg cgctagccga gacgtggcct ttcctgactg cccctgtgtt acctgggcag 900
ctccagatca ctgagccac aatggctgag aaggggtgact gcatcgccag tgtctatggg 960
tatgacctcg gtgggcgctt tgttgacttc caaccctgg gcttcgggtg caatgggtttg 1020
gtgctgtcgg ccgtggacag ccgggcctgc cggaaggtcg ctgtgaagaa gattgccctg 1080
agcgatgcc gcagcatgaa gcacgcgctc cgagagatca agatcattcg gcgcctggac 1140
cacgacaaca tcgtcaaagt gtacgaggtg ctcggtccca agggcactga cctgcagggt 1200
gagctgttca agttcagcgt ggcgtacatc gtccaggagt acatggagac cgacctggca 1260
cgctgctgag agcagggcac gctggcagaa gagcatgcc agetgttcat gtaccagctg 1320
ctccgcgggc tcaagtacat ccactccgcc aacgtgctgc acagggacct gaagcccgcc 1380
aacatcttca tcagcacaga ggacctcgtg ctcaagattg gggatttcgg gttggcaagg 1440
atcgttgatc agcattactc ccacaagggt tatctgtcag aagggttggg aacaaagtgg 1500
taccgttccc cagactgct cctttccccc aataactaca ccaaagccat cgacatgtgg 1560
gocgccggt gcacctggc tgagatgctt acggggagaa tgctctttgc tggggccctg 1620
gagctggagc agatgcaact catcctggag accatccctg taatccggga ggaagacaag 1680
gacgagctgc tcagggtgat gccttccttt gtacgagca cctgggaggt gaagaggcct 1740
ctgcgcaagc tgctccctga agtgaacagt gaagccatcg actttctgga gaagatcctg 1800
acctttaacc ccatggatcg cctaacagct gagatggggc tgcaacaccc ctacatgagc 1860
ccatactcgt gccctgagga cgagcccacc tcacaacacc ccttcgcgat tgaggatgag 1920
atcgacgaca tcgtgctgat ggcgcgtaac cagagccagc tgtccaactg ggacacgtgc 1980
agttccaggt accctgtgag cctgtcgtcg gacctggagt ggcggcctga ccggtgccag 2040
gacgccagcg aggtacagcg cgaccgcgc gcgggttcgg cgccactggc tgaggacgtg 2100
caggtggacc cgcgcaagga ctcgcacagc agtccgagc gcttcctaga gcagtgcgc 2160
tcgtccatgg agcgcgcctt cgaggccgac tacgggcgct cctgcgacta caaggtgggg 2220
tcgccgtcct acctggacaa gctgctgtgg cgcgacaaca agccgcacca ctactcggag 2280
cccaagctca tcctggacct gtgcgactgg aagcaggcgg ccggcgcgcc cccacggcc 2340
acggggctgg cggacacggg ggcgcgcgag gacgagccgg ccagcctctt cctggagatc 2400
gcgcagtggg tcaagagcac gcagggcggc ccagagcacg ccagcccgcc cgccgacgac 2460
cccagcgcc gcttgtctgc ctgcgcccc ggccgccccg ccccggtgga cggcgcgcc 2520
agccccagt tcgacctgga cgtgttcac tccgcgccc tgaagctctg caccaagccc 2580
gaggacctgc cggacaataa actgggcgac ctcaatggtg cgtgcatccc cgagcacct 2640
ggcgacctcg tgcagaccga ggccttctcc aaagaaaggt ggtgagggcg gaggggccgc 2700
tcaggcccc acagagcagg agacccccag agaaagccgg ggctggcagg aggcggccgc 2760
ctccgcccc tctgctgcct tggggttgcc agaacacgtg aaggatccga ggagcgagag 2820
gaatgtccat ttcttaaaact gccttaataa ctagccttta acctgtggga gcgggtttga 2880
acaggacctt ggcttagggg ttgatcactt tcctagcaaa ggggagacca catgtgtgtc 2940
acagggaaga aacggcttta gacagcagtc tgcgggcccc acctgggtgg caggatgccg 3000
agaaatcttg cagaggtagc tccgaaacca tctggcccaa ctagcctcaa ctgacagctg 3060
aggaaaggca attaggccca gagaggcaga gacactcgct taagatcaca ggcttagtgt 3120
gaggacgagc ttgaaatccc agtctcctgg ccccgaggcc agggctctgt caccatagaa 3180
tgtcttcctc tactggggtc gttctggctt tttgttagaa acttggtctg agatgtttct 3240
tcctgttcca ttaccattcg atgttctttt attcagagca atgtttcttg tattctgaaa 3300
ctggaaactg aaccagtttg tttctcctag tcaccaagca tactttctg gctccccag 3360
tacttaaatg ttctcatctg tcgcacccct gtatttgct caccctgca tggtcggaaa 3420
cttctgttcc aggtcagaac agcctggggc ctgtgggtaa aatcagccct tctcccaggc 3480
ctgtgcacac accccctcag cactccctat gcactttcct gacacgcaaa gacacagccc 3540
tctttcccca ctgggcgtcc taccacagtg aggttgaagg caccaattcc aagaatccct 3600
ccaacctccc tgccagcact ccccttcac ccacacccg gccccccac ctaaccacag 3660
cgctcttcca gacctacctc gggacctaat gttctctaca tgaactgtc atttgaggga 3720
cagcagtggg gtcctgccat agagcaaagtg tgttaggaga gaaggtttca catgggaccc 3780
aacatccctc atcaatactt tcctgagttt gatcatccat ttagccttga caaacagcag 3840
accctacaga gatgtgttgg agagcacgtc gtgaccttgg gggcaaggaa tccagaaagg 3900
taggaagata tgaaaagaga ggtgtcaaca gcaagggtc ttaggggtca ggcaccagca 3960
tgagagacct atgacaaagg aagggactca aagcagcaat gccctcata gtgtaggcta 4020
aggtgagttt ggtgcatgca aacctgtgtg ctcaccaca gagcatgggg taatgggtgtg 4080
tagacacagg ccctctgcag aagcgtgggg tggggacact gacagccct atctggctcc 4140

```



```

caggaacatt ctaccatttc tgcactgggt gttcagctcc ttctcttccc ccaacactcc 4200
caaagatacc cacaggaagt ccagccagtt tccaggtaga ggattccacc agttgggtctt 4260
gggctgcgtt caccctcaca tcacagcacc ttaaactctaa tcagcaaact ataatttgta 4320
cgttgaaacc tgcaacacat tagaaactta tatttataaaa cagaattaat cacactgacc 4380
aacttttataa tggaaaatat gtaaataagga agtggttggg ttttggtttt tctttaagaa 4440
aaagaaatgt acaccactcc tcatgtgcc a tttgtcctc agagggcggc tttacttttt 4500
tggtaaagga acaagctgct ggccttgacc aggagttcat atataactgt tattacagag 4560
gaattgttat aactactaat gtttttataa aatttattaa acattattaa acttgatcag 4620
gtcaggccaa ataaagtttt attggaacac aaaaaaaaaa aaaaaaaaaa aaaaaa 4675

```

&lt;210&gt; 18

&lt;211&gt; 4407

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3736769CB1

&lt;400&gt; 18

```

gacgaatcca aacctggagg ggcgtgcgct ccgtaaggct ggtcggcttc ttccgcgcgt 60
tccgctttgc ttcactgctt tctcccccatt tcttgggcct gtggacctgt cctgaggcag 120
aggccgagat gcgcgcaacc gcgggagcag ccaagtggac tggactcttt tcttgactta 180
gctaccagga gctagagatg ctgttattct atcgtatgtg agaagtcggc ccagagatgg 240
aaaactttat tctgtatgag gagatcggaa gaggaagcaa gactgttgtc tataaagggc 300
gacggaaggg aacaatcaat tttgtagcca ttctttgtac tgataagtgc agaaggcctg 360
aaataaccaa ctgggtccgt ctcaccctgt aaataaaaca caagaatatt gtaacttttc 420
atgaatggtg tgaaacaagc aaccacctct ggctagtggg ggaactctgc acaggtggtt 480
ccttaaaaac agttattgct caagatgaaa acctcccaga agatgttgtg agagaatttg 540
gaattgacct gattagtgga ttacatcatc ttcataaact tggcattctc ttttgtgaca 600
tttctcctag gaagatactc ttggaagggc ctggcacact gaagtttagc aacttttgct 660
tggcaaaagt ggaagtgtaa aatttgggag agttctttgc tttggtggca gcagaggag 720
gaggaggtga taatggggaa aatgtcctga agaaaagcat gaaaagtaga gtcaaaggat 780
ctcctgtata tacagacca gaagttgtga ggggtgctga cttttccatc tccagtgaac 840
tctggtcttt gggctgtctg ctttatgaaa tgttttcagg aaaacctcca ttcttctcag 900
aaagtgtttc agaattaact gaaaagatct tatgtgaaga tcctttgcca cctattccga 960
aagattcttc tcgtcctaaa gcttcttcag attttattaa tttgcttgat gggttacttc 1020
aaagagatcc tcagaaaaga ttgacttggg caaggctact gcagattca ttttgggaga 1080
aagcttttgc tggagcagat caggaatcaa gcgtcgaaga tctcagctc agcagaaaca 1140
ctatggagtg ttctgggcca caagattcca aggagctttt gcagaactct cagagtagac 1200
aagcaaaagg gcacaagagt ggtcaaccac taggtcactc tttcagacta gaaaatccaa 1260
ctgagtttgc gcctaagagt actcttgagg gtcaattgaa tgaatccatg tttcttctca 1320
gttctcgctc tactcccaga actagcactg cagtggaggt aagtcctggg gaggatatga 1380
ctcactgttc accacagaag acttctcctc tgaccaagat tacaagtgga cacctgagtc 1440
agcaggacct ggaatcccag atgagagagc ttatctacac ggactcagat cttgttgatca 1500
ccccattat cgacaatcca aagataatga aacagccacc agttaaattt gatgcaaaaa 1560
tattgcatct accaacaatat tcagtggata agttattatt tctgaaagat caagattgga 1620
atgacttttt gcaacaagtg tgctcgaga tcgactccac tgagaagagc atgggggcct 1680
cccagacca gctgaatctc ctttgcattt tgtgcgtggg ggctggtcac caggaggtgg 1740
ccaccaggct cctccattcc cccctgttcc aattgctaatt ccagcatttg cggatagctc 1800
caaactggga tatacgggac aaggttgctc acgtgattgg tttactggct tcgcacacaa 1860
ctgagctcca ggaaaatata cctgttggtg aggcaattgt tctcttaact gaattaatta 1920
gggaaaactt caggaacagc aaattaaaac agtgcccttt accaaccctt ggggagctga 1980
tctatcttgt agccaccag gaagaaaaaa aaaagaacct tagagagtgc tgggctgttc 2040
ccttggtgct atacacagtg ctaatgaggt gccttcggga aggggaagag cgtgttgatga 2100
atcacatggc agcaaaaatt attgaaaatg tctgtaccac ctttctgct cagtcccagg 2160

```

```

gctttattac aggagaaata ggaccatttt tgtggtacct attcagacac tccactgctg 2220
attctcttag gataacagca gtatcggcct tgtgtagaat cactcgccat tctcctactg 2280
ccttccagaa tgttattgaa aagggtgggac tgaactcagt aataaactcc ctggcctctg 2340
ccatctgcaa agttcagcag tacatgttga ccttattcgc tgccatgttg tccctgtggga 2400
ttcatcttca aagactaatc caagaaaagg gttttgtctc cacaattatc cgtttacttg 2460
acagccctc aacatgcatt agagcaaaag ccttcctggt tcttctatat attttgattt 2520
ataaccgtga gatgttgctg ctgagttgcc aagcaagact ggtgatgtac atcgagagag 2580
acagcagaaa gaccactcca ggcaaggagc agcaaagtgg caatgaatac ctgtccaaat 2640
gcctggatct tctcatctgt cacattgtgc aggagctgcc acgaatcctg ggtgacattc 2700
ttaactcctt ggctaagtgt tctggacgta aacacccatc aacagttcaa gtgaaacagc 2760
tgaagtgtg tctccccctg atgcctgtag tgcttcacct cgtaacttca caggatattc 2820
gacctcaagt tgtgacagaa gattttcttt tcagctatgg aactattctt agtcatatta 2880
aatctgtaga ctcaggagaa acgaacatag atggagccat aggactgaca gcatcagaag 2940
aatttatcaa gatcacattg tcagcttttg aagcaataat acagtatcct attttattga 3000
aagactatcg ctccacggtt gttgactata tactgccacc cttggtgtcc ttggttcaaa 3060
gccaaaatgt ggagtggaga ctcttttagt tgcggttgct ctcagaaacc acatctctac 3120
tcgtgaacca ggagtgtggg gatggcaagg agaaggccag tgttgattct gacagcaatc 3180
ttctggctct cattcgagat gtcttacttc cccagtatga gcacattctt ttagaacctg 3240
accagtagc agcatatgct ctgaaactgc tagtcgcgat gactgaacac aaccaactt 3300
tcacaagact tgtggaagaa agcaaactga tccactcat ttttgaagta actctggaac 3360
atcaggagag cattctgggt aataccatgc aaagtgtgat tgcattactc agcaatctag 3420
ttgcctgcaa agattcgaat atggaaactac tttatgaaca aggacttgtc agtcacatct 3480
gtaacctgct cactgaaact gccacactgt gcttggatgt ggacaataaa aacaacaatg 3540
agatggcagc tccactgctc ttttccctgc ttgatatttt gcacagcatg ctgacctata 3600
cctccggtat tgtacggctg gctttgcagg cccagaagtc tggctcagga gaggaccctc 3660
aggctgcaga agacctgctg ctgctcaaca gacctctgac agacctgatt agcctgctca 3720
ttccactgct tccaatgaa gatcctgaga tttttgatgt ttcattccaag tgcctgtcta 3780
tactggttca gctgtatgga ggggaaaacc cggacagcct ctctcctgaa aatgtggaaa 3840
tttttgccta tttactgaca tccaaggagg acccaaagga gcagaagcct ctgttaagga 3900
ttctcagaag aatgatcacc tccaatgaga agcacttggg gagcctcaag aatgcaggca 3960
gcctcctgcy ggctctggag cggctggccc ctgggagtgg ttcatttgcc gacagtgcgg 4020
tggctccctt ggccctggaa atcctccaag ccgttgggca ctaggcaaga aggtgcttag 4080
cacaagcccc cctgtggcc ccagccctcg gatgcataag caaggtcagc tcccagacac 4140
ctttgccaca tccctcaca gctgtctttg gacctataa agtcagctta acccagaacc 4200
tgggtgcccc agtgctcact aaccccaggg cctagaaaac tgactcagaa tggacttcct 4260
tggttcctgt ggaatgcac tgggaagccc aggtttgtta gctgttctca gaaatgttct 4320
ttccctctct gtgtgggcca ggtgggctaa ggttagcact gcctgtggta ataaagcagt 4380
ggatgcaaag cacaaaaaaa aaaaaaa 4407

```

&lt;210&gt; 19

&lt;211&gt; 4795

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7474632CB1

&lt;400&gt; 19

```

tacgcgcccc cgcccgagca cgcaccgcgc gcagaccttc tgcaacaat gctccggccg 60
ggcggctggc ggctcggaga ccgacggagg ggccggggga gcgcacccag aggaagccgc 120
tctgtgacct acctctgaat cccacaaaac tgcaccagag agccggggcg aagatgaacc 180
agcaccctgt cggctcaaga tgcaccagac cctctgctg aaccccgaga gcctgaaaat 240
gtctgcgtgc agtgactttg tggagacat ctggaacccc gggctctgca agaactgctt 300
ctgcctgcgg agcgaccacc agctgggtggc cggccctccc cagcccagag cgggcagcct 360
gccccctcca ccgcgcctgc ctcccaggcc tgagaactgc cgcttggaag atgaagggtg 420

```

gaacagctca cettactcca agcccacaat tgccgtgaag cccaccatga tgagctccga 480  
 ggcctctgat gtgtggacag aggccaacct gagtgccgaa gtctcgagg tcatctggag 540  
 acgagcccct ggcaagctcc ccctcccga gaggaggat gccccgctcg tctacctggg 600  
 cagcttccga ggtgtacaga agcctgctgg tccctctacc tcccctgatg gcaattctcg 660  
 ctgtcccca gcttacacca tggctggcct gcacaacctt gagccccgag gcgagaggaa 720  
 cattgccttc caccgggtga gcttcccga ggagaaggct gtgcacaaag aaaaaccctc 780  
 atttccttac caagaccggc cctccacca ggagagcttc cgccagaaac tggctgcctt 840  
 tgctgggacc acatctggct gtcaccaggg ccctggggcc ctgagggaat ccctgccctc 900  
 ggaggatgac agtgatcaaa ggtgctcgcc ctccggggac agcgagggtg gagagtactg 960  
 ctccatcctg gactgctgcc ctgggagccc tgttgccaag gctgcctccc agactgcagg 1020  
 ttccgggggc aggcaggtg gcagggactg ctaccacag tgctgggagc aggggaagtg 1080  
 ttccggggcc gcagagcagg agaagcggg cccgagcttc ccaaggagt gctgtagcca 1140  
 gggcccccact gcccacccat cctgcctggg ccccaagaaa ctgtccctca cctcggaggc 1200  
 tgccatttct tccgacggcc tctcttggg cagcggcagc ggagcgga gcggcgccag 1260  
 tagcccttc gtcccccacc tcgagagtga ttactgctcc ctcatgaagg aacctgcccc 1320  
 agagaagcag caggacctg gctgccagg ggtgacctt agcagatgcc ttgggctgac 1380  
 gggggagccc cagccccgg cccaacccca ggaggctaca cagcctgaac ccatctatgc 1440  
 tgagagcacc aagaggaaga aggcagctcc ggtgccttcc aagtcacagg ccaagataga 1500  
 acatgcagct gctgccagg gccaaaggcca ggtatgcaca ggtaatgcct gggcccagaa 1560  
 agcagcatct ggctggggcc gggacagccc agaccacact ccccagggtg cagccaccat 1620  
 cacagtcatt gcggcccacc cggaagagga ccacggagc atctacctga gcagccctga 1680  
 ctctgcagtg ggggtgcagt ggccacgagg gcctgtgagc cagaactccg aggtaggtga 1740  
 agaggagact tcggctgggc aggggctgag ctccaggga agccatgctc acagtgccag 1800  
 cgagagcaag ccaaggaga ggcccgccat tcccccaag ttgtccaaga gtagccctgt 1860  
 agggccccg gtgtcacctg ctgctggagg gccccagtg tcaccgctgg ctgaccttag 1920  
 tgatgggagc tctggcggca gcagcattgg gccccagcct ccacccaag gtccctgctga 1980  
 cccgcctcct tcctgcggga ccaacgggtg cgctatcagt gaccatcca ggtgtcccca 2040  
 gcctgcggcc tcgtcagcct cggaacagag gcggcccagg ttccaggcag gcacctggag 2100  
 tcgtcagtc cgcatagagg aagaagagga ggtggagcag gaattgctga gtcacagctg 2160  
 gggaagagag accaaaaatg gcccacgga ccattcaaac tccacgacct ggaccgtct 2220  
 ccacccaca gatggctcct ctgggcagaa cagcaagttt gggaccggga tgagcaaatc 2280  
 cgctctttt gcctttgagt tcccaagga cagaagtggg attgagacat tctcacctcc 2340  
 tcctccgct ccaaagtgc gcacacctt aaaaatgaac aagagcagct ctgatttga 2400  
 aaaagtgagc cagggtctg cagaaagcct cagcccatcc ttcagggtg tccagtcag 2460  
 cttcaccacc ggctccacgg acagcctggc ctcagactct aggacctgca gcgatggagg 2520  
 tccctcgtct gagctggctc actcggccac caacagcggg aagaagctct ttgctcccg 2580  
 tccgtttcct tcaggctcca ctgaggacgt gtccccagtg gggccccagc agccccctcc 2640  
 actccccag aaaaagatag tgagccgggc agcctcttca ccggatggct tcttctggac 2700  
 ccaaggctcc ccaagcccg gaacagcaag cccaagctg aacctagcc actcggaaac 2760  
 caacgtccac gacgaatctc acttttagcta ttctgtgagc cccgggaacc gccaccatcc 2820  
 tgtcttctcc tcttcggatc ctctggagaa agctttcaaa ggcagtggtc actggcttcc 2880  
 ggcagcaggg ctggcgggca acagaggcgg ctgagggagc cctggcctcc agtgcaaagg 2940  
 ggccccctcc gctcatcct cccagctgag cgtgtccagt caagcctcca cggggagcac 3000  
 ccagcttcag ctgcacggtc tcctgagcaa catcagcagc aaggagggca cctatgcaa 3060  
 gctgggggga ctctacacc agtccttggc ccgcttgta gccaatgtg aggacctct 3120  
 catggggcgc cagaaaaagg agctccactt caatgagaat aactggtcgc tcttcaagct 3180  
 gacttgtaac aagccctgct gtgactcggg ggatgccatt tattactgtg ccacctgctc 3240  
 tgaggacccc ggcagcacct atgtgtgaa aatctgcaaa gccccgagc ccaaacagct 3300  
 ctctactgc agcccgctcg tgcccgta ctttaacatc cagcaggact gcggccactt 3360  
 cgtcgcctcg gtgcccgtcca gcatgctcag ctccccgac gcgccaagg accctgtgcc 3420  
 tgccctgccc acacaccccc ctgcccagga gcaggactgc gtggtggtca tcacccgaga 3480  
 ggtgccacat cagaccgct ccgacttctg gcgggactcg gcggccagcc accaggcgga 3540  
 gcccagggcg tacgagcggc gcgtgtgctt cctgcttctg caactctgca acgggctgga 3600  
 gcacctgaag gagcacggga tcatccaccg ggacctgtgc ctggagaacc tgctgctggt 3660  
 gcaactgacc ctccaggccg gcccggggc cggcccgcc cccggcccg cctccgcccc 3720  
 cgccgcccgc gcgcctccct gctcctctgc cggcccgct gctggtggca ctctcagccc 3780

cgagccggc cccgcctccc cggaagggcc cggggagaag cagctgcccc ggctcatcat 3840  
cagcaacttt ttgaaggcca agcagaagcc gggcggcacc ccaaacctgc agcagaagaa 3900  
gagccaggcc cggtggccc ccgagatcgt gtctgcttcc cagtaccgca agttcgatga 3960  
gttccagaca ggcatcctca tctaccagct gctgcaccaa cccaacctgt tcgaggtgcg 4020  
cgcccagctg cgggagagag actaccggca ggaggacctg ccgcccgtgc ccgcgctgtc 4080  
cctctactca cccggcctgc agcagctggc acatctgcta ctggaggccg accccatcaa 4140  
gcgtatccgc atcggcgagg ccaagcgctg gctgcagtgc ctgctgtggg ggccctcggcg 4200  
cgagctggtg cagcagccgg gcacctcgga ggaggcgctg tgcggcacgc tgcacaactg 4260  
gatcgacatg aagcggggccc tgatgatgat gaagtttgcg gagaaggcgg tggatcgag 4320  
gcgggcgctg gagctggagg actggctttg ctgccagtac ctggcgctctg cggagcccg 4380  
ggccctctta cagtcgctga agctcctgca gcttctgtga gccaaagccc agcctgcacc 4440  
gtcgctgccc ctccctgcc taacctttc ctgtctcgcc ttggaagcac ccatgtctcc 4500  
ctgggaaatg gtacagatga ctgggatacc tggatgtaaa atatataaat atatataatg 4560  
aaaatacata taccatatat aaatatgaaa gactaaggat gctgttgcgc gtccacactc 4620  
gtctcctctc tgcactaagt cctccctgtt ttcttctgta attatacaca tttccagttc 4680  
catgcaacgt cctgaggaca gttctgtgaa ctgaatgcag cctggacact ggccctcaata 4740  
cctgttttag gatttcttca cccttttgtc aaattgttat ttaaagaaaa aaaaa 4795

<210> 20

<211> 2386

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7472696CB1

<400> 20

atgggaccgg gtgccctgga gcagggggcg gtgctcggtg gggaggctca ggctgcgcgg 60  
gagcccacgg cggcgcggtt aacgcggggg agacttaggc atggtgggct gcagggtccc 120  
agccctaccc cgcaagaagg tagctaaggc ccggtgagaa atcgagcaca gcgcgggtgg 180  
gccagcagt ctggggaacc cagcacaccc tccgcagctg ttggcctggg tgctaagccc 240  
ctcactacct ggggcaggta gggccggccg gcgctccga gtgcggggcc gccaaaccca 300  
cgcccacccg gaactctagc tggcccgcaa gcgctgtgcg cagccccggt tcccacctgt 360  
gcctctccct ccacacctcc ccgcaagccg agggagctgg ctccggcctc agccagccca 420  
gagaagggct cctcaagcgt agccagaatg ggcgcgagg ccaaggaggc agggagagca 480  
agcaggggct gccagcatgc ttcaccggga agggccgcgc ccccgccgcg ggctgctggc 540  
ccgggtgacg cttccgcctg ctataagagc agcggccctc ggtgcctcct tctgacctc 600  
gcaccagct cggagcccg agcgtgcctc ggcggcctgt cgggtaaggc cgggcctcgg 660  
agcgtgtggc cgggggcccg cgcggggaga ggccaggctc cctcgctcag cttccagccc 720  
ggcaccccc acccccccag ctcaggcccc tcagacccgc agctccctgc gctcgccctc 780  
cccgccagct tcgcgcccc atcccttccg ggcgccccga cggagacaga cgacctctga 840  
tcccccgccc ccgcctgggt acaggccggg ccagcccgcg agggaggag ggagggaatt 900  
gcccttctgt ttctctcacc ttctagttga catcccggt cctccggccc catttgatca 960  
tcgtattgtg acagccaagc aaggagcggc caacagcttc tatactgtga gcaagacaga 1020  
aatcctagga ggaggcggt tcggccaggc tcacaagtgt gaggagacgg ccacagggtc 1080  
gaagctggca gccaaaatca tcaagaccag aggcataag gacaaggagg aggtgaagaa 1140  
cgagatcagc gtcataaacc agctggacca cgcaacctc atccagctgt acgatgcctt 1200  
cgagtcctag aacgacattg tcttggtcat ggagtatgt gatggtggg agctgtttga 1260  
ccgcatcatc gatgagagct acaatttgac ggagcttgat accatcctgt tcatgaagca 1320  
gatatgtgag gggataaggc acatgcatca gatgtacatt ctccacttgg acctgaagcc 1380  
tgagaatatc ctgtgtgtga atcgggatgc tgagcaaata aaaattattg attttggatt 1440  
ggccagaaga tacaaccca gggagaagct gaaggtgaac tttggaaccc cagaatttct 1500  
cgccctgaa gttgtgaact atgattttgt ttcatttccc actgacatgt ggagtgtggg 1560  
ggtcatcgcc tatatgctac ttagcggttt gtgccttcc ctgggtgaca atgatgctga 1620  
gacgctgaac aacatcctgg cctgcagggt ggacttagag gatgaagaat ttcaggacat 1680

```

ctcggaggag gccaggagt tcatctctaa gcttctgatt aaggagaaga gttggcgaat 1740
aagtgcagac gaagctctca agcaccctcg gttgtcagac cacaagctcc actccagact 1800
caatgccagc gtgaccacgg cttcttctc ttcctctttt tctcctgtct gcctgtcttt 1860
tgaagatcag atgctggagt catcttaacc ttaaataaga ttcttttctca ttctttttct 1920
cacagactgc aaatcctgtg cagtattaaa ttacttaaaa tgctttttta aaaaaaattt 1980
tttttgaggc aaaatttcac tctcgttgcc caaactagag tgcaatggca cgatctcagc 2040
tcaatgcaac gtccacctcc ctggttcaag cgattctcct gcctcagtct cacaggtagc 2100
tgggattaca ggctgtgccc accatgcctg gctaattttg tagttttagt ggaaacagag 2160
tttcaccatg ttggtcaggc tgggtctcaa cccctgacct cgtgatccac ccaccttggc 2220
ctcccatagt gctgggatta caggcctgag ccactgcacc cggccttaaa atgctttttt 2280
gagcagttta catttttaaaa ataatgtatt tggtgtttta cagagtattt tcagcataac 2340
tgtttattat taaattattg tatccattac aaaaaaaaaa aaaaaa 2386

```

<210> 21

<211> 3269

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7472343CB1

<220>

<221> unsure

<222> 304, 321, 323

<223> a, t, c, g, or other

<400> 21

```

gcttcttctc catgtccatt tctcggggac tgcccttcca tcccagagcc ctaggcccgt 60
tcttagacag ctcccgctt tccctcaaac cgtcgggtctc acggctgctt ctccctggcct 120
tttgagtgtc tcatccattg gcctcgccct ctctctccca acctctccca tctgtgccct 180
gtcctggact ctttccctgg cttggggctt ccactcttac ccgctccaac ctatcccctt 240
tcatacagta actttctgac actcatatct gaccctgccc tccctgctc aaagcccttc 300
tgtnctccc cagtgcctc ngnacagaat ccaaactcct tgcctggca ttcagggcct 360
ttacaatct caccacacag tagccacaga ctgggacagg agttttctga acacagacac 420
acacacatca catctcccaa gctcaagaag cccaccttct ctcactcctg ccttatcccc 480
attcctgtat gcccaaggcc cagcattaga cccctctctg tcaacacttc acctgttttg 540
tctttgcaag attccgccac tgggcggggg agggggccca gcctggtacc ccacccccac 600
tccagccagg gctcaggctc ccaacaacag aaccagagcc actcaacagc gctggagccc 660
attcgggtgg gcctggggcc cctcatccca agccaggagg gtttctgggg aggggtgcag 720
cccctggcag actgacagtg tggcctgggg gtttgggggt gccagggaag caggggcca 780
cctcatagga ggagacacga gtgcgggtct ctttcccca ctggggggcc tgcgtgtgtc 840
gcagccaggc gggaggcctg ggcggcagag ccagtggtag aggggcctgg gcagggcggt 900
gtctggcagc agcggcacca tgtccaccat ccagtggag actgactgct acgacatcat 960
cgaggtcttg ggcaaggga ccttcgggga ggtagccaag ggctggcggc ggagcacggg 1020
cgagatggtg gccatcaaga tctcaagaa tgacgcctac cgcaaccgca tcatcaagaa 1080
cgagctgaag ctgctgcaat gcatgcagg cctagaccct gaagaggccc acgtcatccg 1140
cttctttag ttcttccatg acgcccctca gttctacctg gtctttgagc tgcgtggagc 1200
aaaccttttc gagttccaga aggagaacaa cttcgcgccc ctccccgcc gccacatccg 1260
tacagtcacc ctgcagggtc tcacagccct ggcccggctc aaggagctgg ctatcatcca 1320
cgctgatctc aagcctgaga acatcatgct ggtggaccag acccgctgcc ccttcagggt 1380
caagggtgatt gacttcggat ccgccagcat tttcagcgag gtgcgctacg tgaaggagcc 1440
atacatccag tcgcgcttct accgggcccc tgagatcctg ctggggctgc cttctgcga 1500
gaagggtggc gtgtggtccc tgggctgct catggctgag ctgcacctgg gctggcctct 1560
ctaccccgcc aacaacgagt acgaccagg gcgctacatc tgcgaaacct agggcctgcc 1620
caagccacac ctgttgcaac ccgctgcaa ggccaccac ttcttcaagc gcaaccccca 1680

```

ccctgacgct gccaacccct ggcagctcaa gtccctcggt gactacctgg ccgagacgaa 1740  
ggtaagggaa aaggagcgcc gcaagtatat gctcaagtcg ttggaccaga ttgagacagt 1800  
gaatggtggc agtgtggcca gtccgctaac cttccctgac cgggaggcgc tggcggagca 1860  
cgccgacctc aagagcatgg tggagctgat caagcgcatg ctgacctggg agtcacacga 1920  
acgcatcagc ccagtgctg ccctgcgcca ccccttcgtg tccatgcagc agctgcgcag 1980  
tgcccacgag accaccact actaccagct ctoctgcgc agctaccgcc tctcgtctga 2040  
agtggagggg aagccccca cgcccgctgt ggccgcagaa gatgggaccc cctactactg 2100  
tctggctgag gagaaggagg ctgcggttat gggcagtggt gccggcagca gccccttctt 2160  
ccgagaggag aaggcaccag gtatgcaaag agccatcgac cagctggatg acctgagtct 2220  
gcaggaggct gggcatgggc tgtggggtga gacctgcacc aatgcggtct ccgacatgat 2280  
ggtcccccctc aaggcagcca tcaactggcca ccatgtgcc gactcgggcc ctgagcccat 2340  
cctggccttc tacagcagcc gcctggcagg ccgccacaag gcccgcaagc cacctgcggg 2400  
ttccaagtcc gactccaact tcagcaacct cattcggctg agccaggtct cgcctgagga 2460  
tgacaggccc tgccggggca gcagctggga ggaaggagag catctcgggg cctctgctga 2520  
gccactggcc atcctgcagc gagatgagga tgggcccac attgacaaca tgaccatgga 2580  
agctgaggtg agccgggtgc gttcaggata cgattaggtt gggaggaggc tcagcacaca 2640  
ctcaccgtg ctcaggatat gattagtgtg tgaggaggct caacacacac tcacccatgt 2700  
tcaggataca attagggact taggaggctc agcacacacc taataccgtc aagatatgat 2760  
aaggctcagc acttactcag ctacttccag gctgtgacaa aaactcaggg cacagtaatc 2820  
tacttataag aagcttgata aagagcctgg gcaacatagt gagatcccgt ctgcaccaa 2880  
aaattagaaa tattagctgg ttcttggtgg cagtgcacct gtaatcccag actactcagg 2940  
attgctgagg tggtaggga tcaactgagc caggagggtc gaggtgcag tgagctgtca 3000  
tcacatcacg tacagaagg gcaataaatg gccatgtca ctgagtaaga ccctagcaca 3060  
tgctcaccct catcaggagg aggtgacaga ggctcagcag aactaatac actaacactg 3120  
cttggtgat gcccctctct cttccccac agaggccaga ccctgagctc ttcgaccca 3180  
gcagctgtcc tggagaatgg ctgagtgcg cagactgcac cctggagagc gtcagggggc 3240  
cacgggctca ggggctccca ccccgccgc 3269

&lt;210&gt; 22

&lt;211&gt; 4888

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7480783CB1

&lt;400&gt; 22

gaggaatcga cagagggtgac ggtgtcgaga gaaaccggat ttgggatctt cgaattttta 60  
gaaatttttt ggaaacttga atcaaattgg ggctacgtat ggatccccac tatacgtag 120  
gcacttttga tgcaactgaa tcgttcatca taactacttg gcagatttta atctctttct 180  
cacagcttaa gaacgaaagg cttagatcaa gtaacttgca cagggttata gatctaaaga 240  
agggcagagg ggagatctgg aaacaggctg gctgactcca aatttggccc attggcaact 300  
aaaactacta tcaactgcag aacgaaaaga aatacatata caaagaaaag aaaagaagtt 360  
aactttttaa agtttttttc tgaatcagcc ttgcaaataa atcataaacc gggtcacagg 420  
ccatattact gcttgaagca gagcaccgaa ttctgataat tttgctccag gaggaggcaa 480  
gggaattaaa acatggctgg gaggtgagcg gacgttttgt acgaggttg gttgaatttg 540  
cttaagaagg tcgcacgcat ctctgataag gaagagaaaa ggccttgcca aggagtattc 600  
tcaacagttc gttcaacggc ttggcagggt aggggaagta caagagccac cctggcggag 660  
ggcagcgtct tgtcacaatt gacagggtgc ggcgcacact gcttctttct cccgcggtt 720  
tcttagaggg cggggaaagc agcaaatacc cctccttgcg ctctagtcct ccaagcctag 780  
ggcggcgggg agtgagggcc cagcgctctg cggtggtgcc aggttccgcc acgcgggccg 840  
cggccgggaat gcggctcggg ggcgcccggg ccaactcgcg ccggcagctg cttaggtcca 900  
gcggggctgc gggagggggc gagctggcga gccgcgggag gggcggagcc ggcgggccc 960  
ggggggcggg gccacccggt tgctcgcgcg gcgcgcgag gctccgcac ccgtcgcgcg 1020  
ggcgggctct ctctgtgaat cctgggtcgc cgatggggga ggtggagccg gggcccgcg 1080

```

gccccgctgga gcccccgagg cggcccggaag cgcccccgag ccgcccggccg ggaggggatcc 1140
gggtcctgaa gatcgtttat gattacttat ccaggctggg atttgatgat cctgtgcgca 1200
tacaggagga ggctacaaat cctgacctcg gctgtatgat tcgattttat ggtgaaaaac 1260
catgccacat ggatcgtttg gatcgaaatcc tattgtctgg catctataat gtacgcaagg 1320
gaaagaccca gctgcataag tgggctgagc gcctagtgtg cctctgtggg acctgcctta 1380
tcgtttcctc agtgaaggat tgtcaaatcg gaaagatgca cattttgcct ctgggtggtg 1440
gaaagataga agaagtgaag cgacggcaat actcccttgc tttcagctca gcaggagccc 1500
aagctcagac ctatcatgtc agcttcgaga ctttggccga gtaccagcga tggcaacggc 1560
aagcatccaa ggtggtgtcc cagcgaatca gtaccgtgga tctctcgtgt tacagcctcg 1620
aggaggttcc tgagcatctc ttctatagtc aagatattac ctacctcaac ttgcgacaca 1680
acttcatgca gttagaaaga cccggaggcc tcgatacact ctacaaattt tctcaactga 1740
agggcctgaa cttgtcccat aataaacttg ggttgtttcc tatattgtta tgcgagatct 1800
ctaccctgac tgagctcaac ctttcctgta atggatttca tgacctacca agtcaaattg 1860
gcaatctgct aaatcttcaa accctctgtc ttgatggcaa ctttctgact actttacctg 1920
aagaattggg aaatcttcaa cagctttcct ccttgggaat ttccttcaac aactttagtc 1980
aaattcctga ggtttatgag aaactcacta tgttagatag agtggttatg gcaggaaatt 2040
gcctggaagt cctgaactta ggggtgctga ataggatgaa ccatatcaag catgtggatt 2100
taaggatgaa ccatttgaac accatggtta ttgaaatct ggagggaat aaacacatca 2160
cccacgtgga tctgcgggac aaccgactga ctgacttggg tcttagctcc ttatgcagct 2220
tggaacagct gcactgtggg cggaatcagc tgaggagct aacactcagt ggcttttccc 2280
ttcggaccct ctatgccagt tccaacaggc tgacagcagt gaacgtctat ccagtacca 2340
gcctgtcac tttcttgat ctctccgaa acctgctaga gtgtgtccct gactgggct 2400
gtgaagcaaa gaagatagaa gtattagatg tgagctataa tcttctcaca gaggttccc 2460
tgagaattct gtagcttg agtcttagaa aactgatgct gggacacaat catgtgcaaa 2520
accttccaac actggtagag cacatcccc tcgaggtgct ggatcttcag cataatgcac 2580
tcacgaggct gccagacacc ctcttctcca aggccttaaa tctcagatac ttgaatgcat 2640
ctgcaaatag tctggagtct ttaccatccg cctgcaactg agaggagagt ttgagtatgc 2700
tgcaagtgtc ttatctgacc aacaatctcc tgacggatca gtgcatacct gtctggtag 2760
ggcacctgca cctgcgaatc ttgcacctg caaacaatca gttacagacc tttcctgcaa 2820
gcaaactaaa taaattggag caattggagg aactgaacct aagtggcaac aagcttaaaa 2880
ccattccac aaccatagca aactgtaaaa ggctgcacac cttgttgca cactccaaca 2940
acatcagcat tttccagaa atactgcagt tgctcagat ccagtttgta gacctagtt 3000
gcaacgactt gacagaaatc ctgattccag agcctttgct tgctacatta caagacctt 3060
acctgactgg aaatacaaat ctggttctgg aacacaagac actggacata tttagccata 3120
tcacaacctt gaaaattgat cagaaacctt tgccaaccac agattctaca gttacgtcaa 3180
ccttctggag ccatggactg gctgagatgg cagggcagag aaataagctg tgtgtctcag 3240
cacttgctat ggatagcttt gcagaggggg tgggagctgt gtatggcatg tttgatggag 3300
accgaaatga ggagctccc cgctgctgc agtgatcagat ggcagatgtg cttttagaag 3360
aggtagcaga gtcaactaat gacacagttt tcatggctaa caccttcttg gtatctcaca 3420
ggaaattagg aatggctggc cagaagttgg gctcctccgc tctcctgtgc tacatccgcc 3480
ctgacactgc cgatccagca agtagcttta gcttgactgt agccaatgtt ggcagtgcc 3540
aagcagtcct gtgccgaggt gggaaagccag tgccctctc taaagtcttc agcctggagc 3600
aggacccaga ggaggctcaa aggtggaagg accaaaaagc catcatcaca gaggacaaca 3660
aagtgaatgg ggtaacctgc tgtaccggga tgctgggctg tacataacctc tacccttggg 3720
tcctcccaa gcccacata tcttccactc cgctgacct tcaagatgag ttgctgattc 3780
tgggaaacaa agcattgttg gaacacttgt cctacacaga agctgtcaat gctgtacgtc 3840
acgtacaaga ccattagca gctgctaaga agctgtgcac attagcgag agctatggct 3900
gtcaggacaa tgtagggcg atggtagttt atttgaatat tgggaagaa ggctgcactt 3960
gtgaaatgaa tgggctcacc ctcccaggtc ctgtgggatt tgcttcaacc accactatca 4020
aggatgcccc taagccagcc actccatcct ctgacagtgg gattgcctct gattgagcaga 4080
gtgagatgtc cacctcagag gtgagcagtg aagtggggtc cactgcttct gatgagcata 4140
atgctggggg cctggacact gccttgcttc cgaggccaga gcggcgctgc agcctccacc 4200
caacacccac ctctgggctg tttcagcgcc agccttcttc tgctaccttc tccagtaacc 4260
agtctgacaa cggcctggac agtgatgatg accagccgt tgaggggggtc ataaccaatg 4320
gcagcaaggt agagggtgaa gtagacatcc actgctgcag ggggagggat ctggagaact 4380
caccctct catagagagt tctcctaccc tgtgttctga ggaacatgct agagggtcgt 4440

```

gttttgggat ccgaagacag aacagtgtga atagtggcat gctcctgcc atgagcaagg 4500  
acaggatgga gttacagaag tctccctcca cctcctgcct ctatgggaag aaactctcca 4560  
atggctctat tgtgccccta gaggacagcc tgaacctcat tgaagtggcc acagaagtgc 4620  
ccaagaggaa aactggctat tttgctgccc ccactcagat ggaaccagag gaccagtgtg 4680  
ttgtgcctca tgacctgga gaagaagtga aggaacaaat gaaacagcac caggacagcc 4740  
ggctcgagcc tgagcccat gaagaggatc ggaccgagcc cccggaggag ttcgacacag 4800  
cactatgact gcccactgg gcacagtgtg ggaggaggct gtgcagggtt ggggtaggga 4860  
cttgctagag gcattctgcc tctacatt 4888

<210> 23

<211> 2380

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7477063CB1

<400> 23

gctttctggg gagccctgac ggggggggtcc gcccaggggc caccacacat caggaagccc 60  
tgagatggag cgcagggcct ccgagacccc tgaggatggg gaccagagg tgagagacgg 120  
gagccagggg aggaagacag agagagaaaa gaaggagaga gaggcaggag gcaacacggg 180  
gagagagagg aagagaaaca gaaaacaggg aaacgcaaag acctcagaga cccagaccca 240  
gggagagagg atgacagaga tccaaggaca cacagcgaga ccaaggcccc cagagaaatg 300  
gactcagaaa cacacagcgg gatgcacagg ggaagaggac ggtggagaga gagaccaggg 360  
tgggcagggg gactgtgtgg gctgaggatg cacccccact cggggctggg ggtcctctggc 420  
ctgctcccac agacaggggc aggcggggcc tctgtggcag tgaccctaa cctctccagg 480  
acacaaaaac aggtggccc agtccggggg gacacagcca cagccctcca acggctgggtg 540  
gagctgacga ccagcagggt gaccccggtg aggagccttc gggaccagta ccacctcatc 600  
cggaagctgg gctccggctc ctacggccgc gtgtcctcgc cccagcctca ccagggggggt 660  
ccagctgtgg ctctgaagct cctgcgtcgc gatttgggtc tgagaagcac ctctctgagg 720  
gagttctgtg tgggcccgtg cgtctctgca caccaggcc tgctgcagac cctggcagga 780  
ccctctacaga cccccgcta ttttgccttc gcccaggagt acgcgccctg tggggacctc 840  
agcgggatgc tgcaggaaag gggectccca gaactgctgg tgaagcgggt ggtggcccag 900  
ttggcaggag ctctggactt cctccacagc cgggggctgg tccacgcaga tgtcaaact 960  
gacaacgtgc tggttcttga cccggtctgc agccgtgtgg ccctgggaga cctgggtctg 1020  
acccggccag agggcagccc gaccccgcc ccaccagtgc ctctgcccac ggcaccgcct 1080  
gagctctgtc tctgtctacc gcccgacacc ctgcctctgc ggccagccgt ggactcctgg 1140  
ggcctggggg tgttctctct ctgtgctgcc actgcctgtt tcccttgagg cgtggcactg 1200  
gcccccaacc ctgagttcga ggccttcgct ggctgggtga ccaccaagcc tcagccacct 1260  
cagccaccac caccctggga ccagtttgcg ccccagccc tggccttget ccaggggcctt 1320  
ctggacctgg atcccagac taggagcccc ccactggctg tcttgactt cctgggggac 1380  
gactgggggt tgcaaggga cagagaggga cctgggggtt tggggagcgc cgtgtcctat 1440  
gaggacaggg agggaggagg ctcaagcctg gaggagtggg cagatgaggg tgatgacagc 1500  
aaaagtgggt ggaggacggg gacagatggg ggagctccct gaccaggtga cagagacagg 1560  
tggccagagc ctgggcccaga ggcccttccc ccagcccca ggccacctg gatggagaga 1620  
cagctgcacc cgggaggaca gagagggaac acattgcact ctccctacgg aacgctgggg 1680  
ttggacgcgc attcctcttc caactgagga cccaagagcc cagctcctgc ccctcctctc 1740  
tcagacgcag gatccaggct cccatccctt cctccctcaa ggaccttcc tcttccctcc 1800  
gtccagagtg tcttcttttt tatttttatt ttttttagac agaattctcg tctgtgacct 1860  
aggctggagt gcagtgggc gatctcagct cactgcaacc tccccctccc aggttcaagt 1920  
gattctcctg cttcaacctc ccaagtagct gggattacac ccgcaccacc acaccgggt 1980  
aattttgtat ttttagtaga ggtgggggtt caccatgttg gccaggctgg tctcgaactc 2040  
ctgacctcaa atgatccacc cgtcttgccc tcccaaagtg ctgggatgac aggcattgagc 2100  
caccacgcct ggccatcatg gagttctgat gggatcactc ctctcctcta agaccttccg 2160  
tggtcccgcc tgtccatgtc acaagatctg aacttttagga tccgacacca tataaacctc 2220



atcctccact ctcccatctt tctggggctg tgggtgatgt catggggcgc ccactcatct 2280  
 gttctgggac ctaaaatcta cacatgggtg agaacacaca cagcacgcg cacacacaca 2340  
 tgggtgcata cggtgtgcac acacacgcac ataacaggtg 2380

<210> 24  
 <211> 3111  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7475394CB1

<400> 24  
 atggccttgc ggggcgcgcg gggagcgacc gacaccccg tgcctcggc cgggggagcc 60  
 cccggcggct cagcgctctc gtcgtccacc tcctcgggcg gctcggcctc ggccggcgcg 120  
 gggctgtggg ccgcgctcta tgactacgag gctcggcgcg aggacgagct gagcctgcgg 180  
 cgccggccagc tgggtggaggt gctgtcgcag gacgcgcgcg tgcggggcga cgagggtctg 240  
 tgggcaggcc aggtgcagcg gcgcctcggc atcttccccg ccaactacgt ggctccctgc 300  
 cgcccgcccg ccagcccgcc gcgcgcgcgc tcggggccca gctccccggt acacgtcgcc 360  
 ttcgagcggc tggagctgaa ggagctcatc ggcgctgggg gcttcgggca ggtgtaccgc 420  
 gccacctggc agggccagga ggtggccgtg aaggcggcgc gccaggaccg ggagcaggac 480  
 gcggcgcccg ctgccgagag cgtgcggcgc gaggctcggc tcttcgccat gctgcggcac 540  
 cccaacatca tcgagctgcg cggcgtgtgc ctgcagcagc cgcacctctg cctggtgctg 600  
 gagttcggcc gcggcggagc gctcaaccga gcgctggccg ctgccaacgc cgccccggac 660  
 ccgcgcgcgc ccggcccccg ccgcgcgcgc cgcctccctc cgcacgtgct ggtcaactgg 720  
 gccgtgcaga tagcgcgggg catgctctac ctgcatgagg aggccttcgt gccatcctg 780  
 caccgggacc tcaagtccag caacattttg ctacttgaga agatagaaca tgatgacatc 840  
 tgcaataaaa ctttgaagat tacagatttt ggggttggcg gggaatggca caggaccacc 900  
 aaaatgagca cagcaggcac ctatgcctgg atggcccccg aagtgatcaa gtcttccttg 960  
 ttttctaagg gaagcgacat ctggagctat ggagtgtgc tgtgggaact gctcaccgga 1020  
 gaagtccctc atcggggcat tgatggcctc gccgtggctt atggggtagc agtcaataaa 1080  
 ctacttttgc ccattccatc cacctgcctc gagccgtttg ccaagctcat gaaagaatgc 1140  
 tggcaacaag accctcatat tcgtccatcg tttgccttaa ttctgaaca gttgactgct 1200  
 attgaagggg cagtgtgac tgagatgcct caagaatctt ttcattccat gcaagatgac 1260  
 tggaaactag aaattcaaca aatgtttgat gagttgagaa caaaggaaaa ggagctgcga 1320  
 tccccggaag aggagctgac tcgggcggct ctgcagcaga agtctcagga ggagctgcta 1380  
 aagcggcgtg agcagcagct ggcagagcgc gagatcgacg tgctggagcg ggaacttaac 1440  
 attctgatat tccagctaaa ccaggagaag cccaaggtaa agaagaggaa gggcaagttt 1500  
 aagagaagtc gtttaagct caaagatgga catcgatca gtttaccttc agatttccag 1560  
 cacaagataa ccgtgcaggc ctctcccaac ttggacaaac ggcggagcct gaacagcagc 1620  
 agttccagtc ccccgagcag ccccaaatg atgccccgac tccgagccat acagttgact 1680  
 tcagatgaaa gcaataaaac ttggggaagg aacacagtct ttcgacaaga agaatttgag 1740  
 gatgtaaaaa ggaattttta gaaaaagggt tgtacctggg gaccaaattc cattcaaatg 1800  
 aaagatagaa cagattgcaa agaaaaggata agacctctct ccgatggcaa cagtccttg 1860  
 tcaactatct taataaaaaa tcagaaaacc atgcccttg cttcattggt tgtggaccag 1920  
 ccagggtcct gtgaagagcc aaaactttcc cctgatggat tagaacacag aaaacaaaaa 1980  
 caaataaaat tgcctagtca ggcctacatt gatctacctc ttgggaaaga tgctcagaga 2040  
 gagaatcctg cagaagctga aagctgggag gaggcagcct ctgcgaatgc tgccacagtc 2100  
 tccattgaga tgactcctac gaatagtctg agtagatccc ccagagaaa gaaaacggag 2160  
 tcagctctgt atgggtgcac cgtccttctg cctcgggtgg ctctgggact ggacctcaga 2220  
 gagcttcata aagcacaggc tgctgaagaa ccgttgccca aggaagagaa gaagaaacga 2280  
 gagggaaatct tccagcgggc ttccaagtcc cgcagaagcg ccagtccctc cacaagcctg 2340  
 ccatccacct gtggggaggc cagcagccca cctccctgc cactgtcaag tgccctgggc 2400  
 atcctctcca caccttcttt ctccacaaag tgcctgctgc agatggacag tgaagatcca 2460  
 ctgggtggaca gtgcacctgt cacttgtgac tctgagatgc tcaactccga tttttgtccc 2520

actgccccag gaagtgggtcg tgagccagcc ctcatgccaa gacttgacac tgattgtagt 2580  
gtatcaagaa acttgccgtc ttccttccta cagcagacat gtgggaatgt accttactgt 2640  
gcttcttcaa aacatagacc gtcacatcac agacggacca tgtctgatgg aaatccgacc 2700  
ccaactgggtg caactattat ctcagccact ggagcctctg cactgccact ctgcccctca 2760  
cctgtctctc acagtcatct gccaaaggag gtctcaccga agaagcacag cactgtccac 2820  
atcgtgcctc agcgtcgccc tgcctccctg agaagccgct cagatctgcc tcaggcttac 2880  
ccacagacag cagtgtctca gctggcacag actgcctgtg tagtgggtcg cccaggacca 2940  
catccccccc aattcctcgc tgccaaggag agaactaaat cccatgtgcc ttcattactg 3000  
gatgctgacg tggaagggtca gagcaggagac tacactgtgc cactgtgcag aatgaggagc 3060  
aaaaccagcc ggccatctat atatgaactg gagaaagaat tctgtctta a 3111

<210> 25

<211> 1372

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7482884CB1

<400> 25

gacaggctgc tgcataatgct tgtgaagggt atgtactaca taactccaga gagcaccatt 60  
tgcagaggcc acgatgtata caacctacca aacaagacta ggcagtatct gagaagggtc 120  
ctcttagggc agagggcggc aaagtttcat gagccaagtt catctcactg cctgttttta 180  
taaataaagt tttattggaa cacaggcaca ctctgtttgt tattatgttg tttatgacca 240  
ctttcatact accatggcca agttgaatag aaactgcacg gctctcaaag cccaaaatat 300  
tgactatctg accctttaca ataaagcctg ctgaccactt agggagactt cttccagtta 360  
agaaaaaac aaaccattt aattttaaaa ttgcttttac taactgttta agtgtgaaga 420  
gtgttttatt atggtgtttt ttccttttca aagttatctt actcttttgc agtaatgcct 480  
gaaaactcaa actttccata tcggcggtat gaccggctcc ctccaatcca tcaattctcc 540  
atagaaagtg acacggatct ctctgagact gcagagttga ttgaggagta tgagggtttt 600  
gatectacca gacctcgacc aaaaatcatt cttgttatag gtgggtccagg aagtggaaag 660  
ggtactcaga gtttgaaaat tgcagaacga tatggattcc aatacatctt tgtgggagaa 720  
ttattaagaa agaagatcca cagtaccagc agcaatagga aatggagtct tattgccaag 780  
ataattacaa ctggagaatt ggccccacag gaaacaacaa ttacagagat aaaacaaaaa 840  
ttgatgcaaa tacctgatga agagggcatt gttattgatg gatttccaag agatgttgcc 900  
caggctctat cttttgagga ccaaattctgt acccccgatt tgggtggtatt cctggcttgt 960  
gctaatacaga gactcaaaga aagattactg aagcgtgcag aacagcaggg ccgaccagac 1020  
gacaatgtaa aagctaccca aaggagacta atgaacttca agcagaatgc tgctccattg 1080  
gttaaatact tccaggaaaa ggggtctatc atgacatttg atgccgaccg cgatgaggat 1140  
gaggtgttct atgacatcag catggcagtt gacaacaagt tatttccaaa caaaggagct 1200  
gcagcagggt caagtgcct tgatccttcg atgatattgg acactggaga gatcattgat 1260  
acaggatctg attatgaaga tcagggtgat gaccagttaa atgtatttgg agaggacact 1320  
atgggagggt tcatggaaga tttgagaaaa gtgtaaaatt tattttcata at 1372

<210> 26

<211> 1704

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7494121CB1

<400> 26

cacaatgggtc caccacaaaa tgaattatca ggagtgaacc cagaggcacg tatgaatgaa 60

```

agtcctgata cgactggcct gacgggagtc atcattgagc tcggcccca tgacagtcca 120
cagacaagtg aatttaaagg agcaaccgag gaggcacctg cgaaagaaag cccacacaca 180
agtgaattta aaggagcagc ccgggtgtca cctatcagtg aaagtgtgtt agcagcactt 240
tccaagtttg aagttgaaga tgctgaaaat gttgcttcat atgacagcaa gattaagaaa 300
attgtgcatt caattgtatc atcctttgca tttggactat ttggagtttt cctggcttta 360
ctggatgtca ctctcatcct tgccgacctt attttctactg acagcaaact ttatattcct 420
ttggagtata gttctatttc tctagctatt gccttatttt ttctcatgga tgttcttctt 480
cgagtatttg tagaaaggag acagcagtat ttttctgact tatttaacat ttagataact 540
gccattattg tgattcttct gctggttgat gtcgtttaca ttttttttga cattaagtgtg 600
cttaggaata ttccagatg gacacattta ctctgacttc tacgacttat tattctgtta 660
agaatttttc atctgtttca tcaaaaaaga caacttgaaa agctgataag aaggcgggtt 720
tcagaaaaca aaaggcgata cacaagggat ggatttgacc tagacctcac ttacgttaca 780
gaacgtatta ttgctatgtc atttccatct tctggaaggc agtctttcta tagaaatcca 840
atcaaggaag ttgtgcggtt tctagataag aaacatcgaa accactatcg agtctacaat 900
ctatgcagtg aaagagctta tgatcctaag cacttccata atagggtcag tagaatcatg 960
attgatgata ataagtccc cactctacat cagatgggtg ttttcaccaa agaagtaaat 1020
gagtggatgg ctcaagatct tgaaaacatc gtagcgattc actgtaaagg aggcaaagga 1080
agaaccggga ctatggtttg tgccctcctt attgcctcgg aaatatTTTT aactgccgag 1140
gaaagcctat attatttttg agaaaggcga accaataaaa cccacagcaa taaatttcag 1200
ggagtagaaa ctctctctca gaatagatat gttggatatt ttgcacaagt gaaacatctc 1260
tacaactgga atctccctcc aagacggata ctctttataa aaagattcat tatttattcg 1320
attcgtgggtg atgtatgtga tctaaaagtc caagtagtaa tggagaaaaa ggttgtcttt 1380
tccagtactt cattaggaaa ttgttcgata ttgcatgaca ttgaaacaga caaaatatta 1440
attaatgtat atgacgggtc acctctgtat gatgatgtga aagtgcagtt tttctcttcg 1500
aatcttctca aatactatga caattgtcca tttttcttct ggttcaacac gtcttttatt 1560
caaaataaca ggctttgtct accaagaaat gaattggata atccacataa, acaaaaagca 1620
tggaaaattt atccaccaga atttgcgtgtg gagatacttt ttggcgaaat gacttccaat 1680
gatgtttag ctggatccga ctaa 1704

```

&lt;210&gt; 27

&lt;211&gt; 5563

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6793486CB1

&lt;400&gt; 27

```

ggtctctgag ggtcaggaga ggaatcaccc agggatcctc tcccccaaga aacttggccc 60
gcagtgcagtg gccaaggggtg aaagtgcagag tgctgaagag gggcctgtct taccctgctt 120
ctgctcctcc agggagttga taatgcttgt gggcagcaac ggcagggtcca cgctgacctt 180
catcgttttg cccctttgcc ttccagaagg atttgcccct gcccggaaa agcagcaggt 240
gagcctgggg agagacgtgg cgggggctgg ccgccccttc tcccgcggg gtgggggcgc 300
caggcgagtg cgcgggcggg gcctggctcg ccggcacgca gcagcgcccc ctcccgccg 360
gctgcagcgg cagggcctc cccttcctg ccctccctg cccgtgcgg cggtgcagc 420
ctccgccccg cgcgcttgct cccgcgcgg ccgcgcgcgc cgctccgcgc gctgctgcgc 480
cacctgccac catgtgcgg ccgcccgggc atgtctgact ctctctggac cgcgctttct 540
aatttctcga tgccctcctt cccggcgggc agtatgttcc gccgcacca gagctgcgcg 600
accagtaatc ggaaaagcct catcctgacc agcacttcac ccacgtacc gagacccac 660
tccccgctgc caggtcacct aggcagcagt cccctggaca gccccgaaa cttctcccc 720
aacacccccg cccacttctc gtttgcttcc tcccgaaggg cggacggacg ccggtggtct 780
ctggcctcgc tcccttcac tggttatggc accaacacgc ccagttccac cgtctcgtcc 840
tcctgctcct cccaggagcg ccttcaccag ctgccctacc agcccacggt ggacgagctc 900
cacttcctct ccaaacactt cgggagcacc gagagcatca cagacgagga tgggtggcgt 960
cgctccccag ccgtgcggcc ccgctcacgg agcctcagcc ccgggcgctc cccctcctcc 1020

```

tacgacaacg agatcgtgat gatgaatcac gtctacaagg agaggttccc gaaggccact 1080  
 gcgcagatgg aggagaagct gcgcgacttt acccgccctt acgaaccgga cagcgttctg 1140  
 cctctggccg atggcgtgct cagcttcctc caccaccaga tcatcgagct ggcccgggac 1200  
 tgccctgacca agtcccgtga cggcctcctc accacggctt acttctatga attgcaggag 1260  
 aacctggaga agctccttca agacgcctat gaacgctctg agagcttgga ggtggccttc 1320  
 gttactcagc tgggtgaagaa gttgcttatt atcatctcac gccctgcgag gctgctggag 1380  
 tgccctggaat tcaaccccgga ggagttctac cacctgctgg aggcggccga aggacacgcc 1440  
 aaggagggcc accttggtgaa gacggacatc ccccgctaca tcatccgcca gctgggcctc 1500  
 acccgtgacc cctttccaga tgttggtgat ctggaggaaac aggacagtgg tggttccaac 1560  
 acccctgagc aagacgatct ctctgagggc cgcagcagca aggccaaagaa accgcggggg 1620  
 gagaatgact tcgataccat caagctcata agcaacgggtg cctacggcgc tgtctacctg 1680  
 gtgcggcacc gcgacacgcg gcagcgcttt gccatgaaaa agatcaacaa gcagaacttg 1740  
 atcctccgca accagatcca gcaggccttt gtggagcgcg atatcctcac cttcgccgag 1800  
 aaccggtttg tggtcggcat gttctgctcc tttgagactc ggccccaact ctgcatggct 1860  
 atggaatatg tgggaaggcgg cgactgtgcc accctgctga agaattattg agcgtgccc 1920  
 gtagagatgg cccgcatgta ctttgctgag acggtgctag ccctggagta tttgcacaac 1980  
 tatggcatcg tgcaccgcga cctcaagcct gacaacctcc ttatcacctc catgggtcac 2040  
 atcaagctca cagatttcgg cctctccaag atggggctca tgagcctcac caccaactta 2100  
 tatgaaggcc acatcgagaa ggacgcccga gagttcctgg acaaacaggt gtgtgggacc 2160  
 ccagagtaca tcgcgcccga ggtcatcctg cgtcaaggct acggcaagcc agtggactgg 2220  
 tgggctatgg ggatcatcct ctacgagttc ctgggtgggt gtgtgccctt cttcgagac 2280  
 acaccagagg agctatttgg acaggctatc agtgatgaca tcctgtggcc cgagggggat 2340  
 gaggccctac ctacggaggc ccaactcctc atatccagcc tcctgcagac caacctctg 2400  
 gtcaggcttg gggcaggcgg cgcttttgag gtgaagcagc acagtttctt tcgagacctg 2460  
 gactggacag ggctgctgag gcagaaggcc gagttcatcc occacctaga gtcggaagat 2520  
 gacactagct actttgacac ccgctcagac aggtatcacc acgtgaactc ctatgacgag 2580  
 gatgacacga cggaggagga gcccggtgaa atccgccagt tctcttctctg ctctccgcgc 2640  
 ttcagcaagg tgtatagcag catggagcag ctgtcgcagc acgagcccaa gacccagta 2700  
 gcagctgcag ggagcagcaa gcgggagccg agcaccaagg gccccgagga gaaggtggcc 2760  
 ggcaagcggg aggggctggg cggcctgacc ctgctgaga agtccataac cagcccccct 2820  
 cctttgggcg cccgcgcgcg tttctcggcg ctgctggagc ccagccgctt cagcgcccc 2880  
 caagaggacg aggatgaggc ccggtgcgc aggcctcccc ggcccagctc cgaccccgcg 3000  
 ggatccctgg atgcacgggc ccccaaagag gagactcaag gggaaggcac ctccagcgcc 3060  
 ggggactccg agggcactga ccgtccacgc ccaggtgacc tctgcccacc ctccaaggat 3120  
 ggggatgcat caggcccaaag ggctaccaat gacttggttc tgcgcccggc gcggcaccag 3180  
 cagatgtcag gggatgtggc agtagagaag aggccttctc gaactggggg caaagtcatc 3240  
 aatcagcct cagccactgc cttatctgtc atgattcctg cagtggacct acatggaagt 3300  
 tcaccccttg ctagtcccat gtctccacga tctctgtcct ccaacctc ctcacgggac 3360  
 tcctcaccga gccgggacta ctcaccagct gtcagtgggc tccgctcccc catcaccatc 3420  
 cagcgctcgg gcaagaagta tggcttcaca ctgctgcca tccgtgtcta catgggtgac 3480  
 acggatgtct atagtgtcca ccacattgtc tggcatgtgg aggaaggagg cccagcccag 3540  
 gaggcaggac tctgtgctgg ggacctcatc acccagctga atggggagcc tgtgcatggc 3600  
 atggtgcata ctgaggtcgt ggagctgata cttaagagtg gcaacaaggt agcagtgacc 3660  
 acaacgccct tcgaaaaatac ctctatccgc attggtcccg caaggcgag cagctacaag 3720  
 gctaataatg ctcggaggaa caagcgacct tccgccaagg agggccagga gagcaagaag 3780  
 cgcagctccc tcttccggaa gatcacgaag cagtogaacc tgctgcatac tagccgctcg 3840  
 ctgtcgtcgc tgaaccgctc gctgtcatcc agcgatagtc tcccgggctc gcctacgcac 3900  
 gggctgccgg cgcgctcgc caccgacagc taccgctcca cgctgactc gcctaccta 3960  
 ggtattacct cctgcacctg cgcggggacc gagcagacgc ccaactcgcc tgcgtcgtcg 4020  
 gcgtcgacc acattcggcc cagcacgctg cacggactgt cgcaaagct ccatcgccag 4080  
 taccgctctg cgcgatgcaa gtcggccggc aacatccctc tatcgccgct ggcacacag 4140  
 ccgtcccca cgcaggcgtc accgcccga ctgcccggcc acacgcgccc caagagtgc 4200  
 gagccccctc gctcgccgct cctcaagcgc gtgcagtcgg ccgagaagct gggagcctct 4260  
 ttgagtgcgg acaagaaggc cgcgtcgcgc aaacacagcc tcgaggtggg ccacccggat 4320  
 ttccgcaagg acttccatgg cgagctggcg ctgcatagcc ttgccgagtc cgacgggtgag 4380

```

acgccccag tcgagggcct tggcgcgccc cggcaggctcg cegtccgccc cctggggccga 4440
caggagtcac ctttgagcct gggcgcggaac ccgttgctgc cggaggggtgc ctccaggcca 4500
ccagtgtcga gcaaggagaa ggaatccccg gggggcgccc aggcgtgcac cccaccccg 4560
gcgacgaccc ccggtggccg gaccctggag cgggacgtcg gctgcacgcg gcatcagagc 4620
gtgcagacgg aggatggcac tggcgggatg gccagggctg tggccaaggc ggcgctgagc 4680
ccggtgcagg aacacgagac aggccggcgc agcagctctg gcgaggcggg cacaccctg 4740
gtacccattg tcgtagagcc tgcgcggccc ggggctaagg ctgtggtgcc tcagcctctg 4800
ggcgcggaact ccaaggggtt gcaggaaccc gcacccctgg cgccttccgt gcccagggcc 4860
ccccggggcc gggagcgctg ggtgttgag gtggtggagg agcgcaccac gctgagcggg 4920
cctcgctcca agcccgctc cccaagctc tccccggagc cccagacacc ctccctagcc 4980
ccagcgaagt gcagtgcacc cagcagtgc gtgacccag tcccacccgc atccctcttg 5040
ggctcaggca ccaagcctca agtggggctg acctcccggg gccctgctga agctgtgccc 5100
ccagcaggcc tgacaaaaa aggagtgtcc agtcccgac ccccgggacc atagccaagg 5160
gggtcatcgg ccccgcgctg tacagcctcc gtatacatat gtacacatat aaataaagt 5220
cgtccgtgct gcgtgaaaaa aacaaaaaac aaaacaccga acgaacacaa aaccatgaca 5280
cacaccagac atcaccacac acacaacacc ataatgagcc aacagccgga aaaacacaca 5340
cgacggccgc cacaaaccac ccccccccg ggattagccg cgccgcccgc ggcgagccac 5400
caccccccg cggcacacgc ccggccccc acctaccct gacgcgggaa ctgcgcgcc 5460
ccaccaccac acaacaaaac acggcgcgc caccaacacg ctccagccgc acggaccacc 5520
taccgcgcgc cccctccag cgcaccctag cgcgcgccca ccc 5563

```

&lt;210&gt; 28

&lt;211&gt; 1697

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7494178CB1

&lt;400&gt; 28

```

ggaggcagca tgctaaaccg ggtgcgctcg gccgtggcgc acctggtgag ctccgggggc 60
gctccgcctc cgcgccccaa atccccggac ctgcccacgc ccgcctcggc gccgcccgc 120
gccgtccag aagcgcccag gagccctccc gcgaaggctg ggagcgggag cgcgacgccc 180
gcgaaggctg ttgaggctcg agcgagcttc tccagaccga ctttctgca gctgagcccc 240
ggggggctgc gacgcgccga tgaccacgcg ggccgggctg tgcaaagccc cccggacacg 300
ggccgcgcgc tgcctggag cacaggctac gccgaggtca tcaatgctgg caagagtgcg 360
cacaatgagg accaggcttg ctgtgaagtg gtgtatgtgg aaggtcggag gagtgttaca 420
ggagtacctt gggagcctag ccgaggccag ggactctgct tctactactg gggcctattt 480
gatgggcatg cagggggcgg agctgctgaa atggcctcac ggctcctgca tcgccatatc 540
cgagagcagc taaaggacct ggtagagata cttcaggacc cttcgccacc acccctctgc 600
ctcccaacca ctccggggac ccagattcc tccgatccct ctcaattgct tggccctcag 660
tctgtctggt cttcacagaa ggaagtgagc cacgagagcc tggtagtggg ggccattgag 720
aatgccttcc agctcatgga tgagcagatg gcccgggagc ggctggcca ccaagtggag 780
gggggctgct gtgactggt tgtgatctac ctgctaggca aggtgtacgt ggccaatgca 840
ggcagtagca gggccatcat tgtccggaat ggtgaaatca ttccaatgtc cggggagttt 900
accccgagga ctgagcgcca cgtcttcag ctgcttggt tccctgaaacc agagctgcta 960
ggcagtgaat tcacccacct tgagttcccc cgcagagttc tgcccaagga gctggggcag 1020
aggatgttgt accgggacca gaacatgacc ggctgggctt acaaaaagat cgagctggag 1080
gatctcaggt ttcctctggt ctgtggggag ggcaaaaagg ctcggtgat ggccaccatt 1140
ggggtgaccc gaggcttggg agaccacagc cttaaggctc gcagttccac cctgcccac 1200
aagcccttcc tctctgctt ccctgaggta cgagtgtatg acctgacaca atatgagcac 1260
tgccagatg atgtgctagt cctgggaaca gatggcctgt gggatgtcac tactgactgt 1320
gaggtagctg ccactgtgga cagggtgctg tcggcctatg agcctaata ccacagcagg 1380
tatacagctc tggcccaagc tctggtcctg gggggccggg gtaccccccg agaccgtggc 1440
tggcgcttcc ccaacaacaa gctgggttcc ggggatgaca tctctgtctt cgtcatcccc 1500

```

```

ctgggagggc caggcagtta ctctgaggg gctgaacacc atccctccca ctagcctctc 1560
catacttact cctctcacag cccaaattct gaagttgtct ccctgaccct tcttttagtgg 1620
caacttaact gaagaaggga tgtccgctat atccaaaatt acagctattg gcaaataaac 1680
gagatggata aaaaaaa                                     1697

```

&lt;210&gt; 29

&lt;211&gt; 3280

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7096516CB1

&lt;400&gt; 29

```

ccccggacct cgcgcagggg gccccgggac accccctgcg ggccgggtgg aggaggaaga 60
ggaggaggag gaagaagacg tggacaagga ccccttcct acccagaaca cctgcctgcg 120
ctgccgccac ttctctttaa gggagaggaa aagagagcct aggagaacca tggggggctg 180
cgaagtccgg gaatttcttt tgcaatttgg tttctctctg cctctgctga cagcgtggcc 240
aggcgactgc agtcacgtct ccaacaacca agttgtgttg cttgatacaa caactgtact 300
gggagagcta ggatggaaaa catatccatt aaatgggtgg gatgccatca ctgaaatgga 360
tgaacataat agggccattc acacatacca ggtatgtaat gtaatggaac caaaccaaaa 420
caactggctt cgtacaaact ggatctcccg tgatgcagct cagaaaattt atgtggaaat 480
gaaattcaca ctaagggatt gtaacagcat cccatgggtc ttggggactt gcaaagaaac 540
atttaatctg ttttatatgg aatcagatga gtcccacgga attaaattca agccaaacca 600
gtatacaaaag atcgacacaa ttgctgctga tgagagtttt acccagatgg atttgggtga 660
tcgcatcctc aaactcaaca ctgaaattcg tgaggtgggg cctatagaaa ggaaggatt 720
ttatctggct tttcaagaca ttggggcggt cattgccctg gtttcagtcc gtgttttcta 780
caagaaatgc cccttcaact ttctgtaact ggccatgttt cctgatacca ttccaagggt 840
tgattcctcc tctttgggtg aagtacgggg ttcttggtg aagagtgtcg aagagcgtga 900
cactcctaaa ctgtattgtg gagctgatgg agattggctg gttcctcttg gaaggtgcat 960
ctgcagtaca ggatatgaag aaattgaggg ttcttgccat gcttgacagc caggattcta 1020
taaagctttt gctgggaaca caaatgttc taaatgtcct ccacacagtt taacatacat 1080
ggaagcaact tctgtctgtc agtgtgaaaa gggttatttc cgagctgaaa aagaccacc 1140
ttctatggca tgtaccaggc caccttcagc tcctaggaat gtggttttta acatcaatga 1200
aacagccctt attttggaat ggagcccacc aagtgcacac ggaggggagaa aagatctcac 1260
atacagtgtg atctgtaaga aatgtggctt agacaccagc cagtgtgagg actgtggtgg 1320
aggactccgc ttcattccaa gacatacagg cctgatcaac aattccgtga tagtacttga 1380
ctttgtgtct cacgtgaatt acaccttga aatagaagca atgaatggag tttctgagtt 1440
gagtttttct cccaagccat tcacagctat tacagtgcac acggatcaag atgcaccttc 1500
cctgataggt gtggtaaagg aggactgggc atcccaaat agcattgccc tatcatggca 1560
agcacctgct ttttccaatg gagccattct ggactacgag atcaagtact atgagaaga 1620
acatgagcag ctgacctact cttccacaag gtccaaagcc ccagtgatca tcatcacagg 1680
tcttaagcca gccacaaat atgtatttca catccgagtg agaactgcga caggatacag 1740
tggtacagct cagaaatttg aatttgaaac aggagatgaa acttctgaca tggcagcaga 1800
acaaggacag atttctgtga tagccaccgc cgctgttgcc ggattcactc tctctgcat 1860
cctcacttta ttcttcttga tcaactggag atgtcagtg tacataaaag ccaagatgaa 1920
gtcagaagag aagagaagaa accacttaca gaatgggcat ttgcgcttcc cgggaattaa 1980
aacttacatt gatccagata catatgaaga cccatcccta gcagtccatg aatttgcaaa 2040
ggagattgat cctcaagaa ttcgtattga gagagtcatt ggggcagggt aatttgagaa 2100
agtctgtagt gggcgtttga agacaccagg gaaaagagag atcccagttg ccattaaaac 2160
tttgaaagggt ggccacatgg atcgcaaaag aagagatttt ctaagagaag ctagtatcat 2220
gggccagttt gaccatcaa acatattcg cctagaaggg gttgtcacca aaagatcctt 2280
cccgccattt ggggtggagg cgttttgccc cagcttctct agggcagggt ttttaaatag 2340
catccaggcc cgcacccag tgccaggggg aggatctttg cccccagga ttctgtctgg 2400
cagaccagta atgattgtgg tggaatatat ggagaatgga tccctagact cctttttgctg 2460

```

```

gaagcatgat ggccacttca cagtcattcca gttgggtcga atgctccgag gcattgcatc 2520
aggcatgaag tatctttctg atatgggtta tgttcacga gacctagcgg ctcggaatat 2580
actggtcaat agcaacttag tatgcaaagt ttctgatttt ggtctctcca gagtgtctgga 2640
agatgatcca gaagctgctt atacaacaac ggggtggaaa atccccataa ggtggacagc 2700
cccagaagcc atcgccctaca gaaaattctc ctacagcaagc gatgcatgga gctatggcat 2760
tgtcatgtgg gaggtcatgt cctatggaga gagaccttat tgggaaatgt ctaaccaaga 2820
tgtcattctg tccattgaag aagggtacag acttcagct cccatgggct gtccagcatc 2880
tctacaccag ctgatgctcc actgctggca gaaggagaga aatcacagac caaaatttac 2940
tgacattgtc agcttccttg acaactgat ccgaaatccc agtgcccttc acaccctggt 3000
ggaggacatc cttgtaatgc cagagtcccc tgggtaagtt ccggaatc ctttgtttgt 3060
cacagttggt gactggctag attctataaa gatggggcaa tacaagaata acttcgtggc 3120
agcagggttt acaacatttg acctgatttc aagaatgagc attgatgaca ttagaagaat 3180
tggagtcata cttattggac accagagacg aatagtcagc agcatacaga ctttacgttt 3240
acacatgatg cacattcagg agggaggatt tcatgtatga 3280

```

<210> 30

<211> 2781

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7474666CB1

<400> 30

```

gttgggtctgg ggctggccac atggagcccg ggctggagca cgcactgcgc aggaccctt 60
cctggagcag ccttgggggt tctgagcatc aagagatgag ctccctagag caagaaaaca 120
gcagctcatg gccatcacca gctgtgacca gcagctcaga aagaatccgt gggaaacgga 180
gggccaagc cttgagatgg acaaggcaga agtcggtgga ggaaggggag ccaccaggtc 240
agggggaagg tccccggtcc aggccagctg ctgagtcac cgggctggag gccacattcc 300
ccaagaccac acccttggtc caagctgac ctgcccgggt gggcactcca ccaacagggt 360
gggactgcct cccctctgac tgtacagcct cagctgcagg ctccagcaca gatgatgtgg 420
agctggccac ggagttccca gccacagagg cctgggagtg tgagctagaa ggctgctgg 480
aagagaggcc tgccctgtgc ctgtcccgc aggccatt tcccaagctg ggctgggatg 540
acgaactgag gaaaccggc gccagatct acatgcgctt catgcaggag cacacctgct 600
acgatgccat ggcaactagc tccaagctag tcatcttoga caccatgctg gagatcaaga 660
aggccttctt tgctctggtg gccaacggtg tgcgggcagc ccctctatgg gacagcaaga 720
agcagagctt tgggggatg ctgaccatca ctgacttcac cctggtgctg catcgctact 780
acaggtcccc cctggtccag atctatgaga ttgaacaaca taagattgag acctggaggg 840
agatctacct gcaaggctgc ttcaagctc tggctcccat ctctccta at gatagcctgt 900
ttgaagctgt ctacaccctc atcaagaacc ggatccatcg cctgcctggt cttgaccctg 960
tgtcaggcaa cgtactccac atcctcacac acaaacgcct gctcaagttc ctgcacatct 1020
ttggttccct gctgccccgg ccctccttcc tctaccgcac tatccaagat ttgggcatcg 1080
gcacattccg agacttggtc gtggtgctgg agacagcacc catcctgact gcactggaca 1140
tctttgtgga cggcggtgtg tctgactgct ctgtggtcaa cgaatgtggc caggtcgtgg 1200
gcctctattc ccgctttgat gtgattcacc tggctgcccc gcaaacctac aaccacctgg 1260
acatgagtggt gggagaagcc ctgaggcaga ggacactatg tctggaggga gtcctttcct 1320
gccagcccc cagagcttg ggggaagtga tcgacaggat tgctcgggag caggtacaca 1380
ggctggtgct agtggacgag acccagcatc tcttgggctg ggtctccctc tccgacatcc 1440
ttcaggcact ggtgctcagc cctgctggca tcgatgcctt cggggcctga gaagatctga 1500
gtcctcaatc ccaagccacc tgcacacctg gaagccaatg aagggaactg gagaactcag 1560
ccttcatctt cccccacccc catttgctgg ttcagctatg attcaggctt cttcagccct 1620
cccaaatgct ccttgcccta cctgtgctcc cagaagccct cgggcatgcc cagtgcacca 1680
tgggatgatg aaattaagga gaacagctga gtcaagcttg gaggtccctg aaccagaggg 1740
actaggatta ccccagggcc atctgtgctc catgcccgc catccccttg ccgctgact 1800
gggtcggatg gcccagtggt gtttagtcag ggcttctgga ttctcgggtt tctgggctac 1860

```

```
ctatggcttc agccttcagc tcctgggagt ccagctggtt gttcccagca acgtcgccac 1920
tgccctccta ctctccaggc tttgtcattt caaggctgct gaaatgctgc atttcagggg 1980
ccaccatgga gcagccgtta tttatagaac tgctgttgg aggtggggag tcctccctcc 2040
attcttgtec agaaaactcc ttagctctcg cagtgaacca tgttcttagt ctccagggat 2100
ggatggcctt gtatatggac ccctgagaat gagcaattga gaaaacaaaa caaaaggaac 2160
aatccatgaa cttagatttt attggtttca ctcaaatgc tgcagtcatt tgacctgaac 2220
ttgtggcaag agacttggtc tttctaaatt caaagactag aaggaaaatg gataaaaatc 2280
acaagtgcg tttctcttgc aatgtagcgc tattctactg aaatttcttt ctctcttttt 2340
ctttacaaaa tcataaagaa aaaattaatt cttacttat atagtaggta caactcagcc 2400
tacaaactct aatctgcaag aagcataact ttatttttct aacacagaat gtaatttcta 2460
ttaggaaccc cgtttcagca ggtggtagaa attaatctca gtcaattcaa agtctccccc 2520
tgaccttttc ctggggttaa gctcggtcgg gtgggggtag tggctttaag tcatgtaaga 2580
ctctgttcct cggctatcat catccgtcca tgctacgcac cagcctggac atcccctccc 2640
catctggtca tcagtctggt catgcaaggt ctagccaggg ctcccttcaat tccaçaaagc 2700
ctattgggga cctgtggctt ggagcatgtg gaagagtcga gctcatggcc ctgcagacac 2760
acaaggctac aggaagcaca a 2781
```



(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
24 October 2002 (24.10.2002)

PCT

(10) International Publication Number  
**WO 2002/083709 A3**

(51) International Patent Classification<sup>7</sup>: **C12P 21/06**,  
C12Q 1/68, C12N 15/00, 5/00, 1/20, C07K 1/00, C07H  
21/00, 21/04

(21) International Application Number:  
PCT/US2002/010818

(22) International Filing Date: 5 April 2002 (05.04.2002)

(25) Filing Language: English

(26) Publication Language: English

## (30) Priority Data:

|            |                            |    |
|------------|----------------------------|----|
| 60/282,119 | 6 April 2001 (06.04.2001)  | US |
| 60/283,588 | 13 April 2001 (13.04.2001) | US |
| 60/283,759 | 13 April 2001 (13.04.2001) | US |
| 60/285,589 | 20 April 2001 (20.04.2001) | US |
| 60/287,037 | 27 April 2001 (27.04.2001) | US |
| 60/287,036 | 27 April 2001 (27.04.2001) | US |
| 60/288,608 | 4 May 2001 (04.05.2001)    | US |
| 60/288,712 | 4 May 2001 (04.05.2001)    | US |
| 60/289,909 | 9 May 2001 (09.05.2001)    | US |
| 60/292,246 | 17 May 2001 (17.05.2001)   | US |

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

## (72) Inventors; and

(75) Inventors/Applicants (for US only): RECIPON, Shirley, A. [US/US]; 85 Fortuna Avenue, San Francisco, CA 94115 (US). BURRILL, John, D. [US/US]; 2218 Brewster Avenue, Redwood City, CA 94062 (US). MARCUS, Gregory, A. [US/US]; 73 Maple Way, San Carlos, CA 94070 (US). ZINGLER, Kurt, A. [US/US]; 723 Ashbury Street #C, San Francisco, CA 94117 (US). TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). THORNTON, Michael [US/US]; 9 Medway Road, Woodside, CA 94062 (US). BOROWSKY, Mark, L. [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). BURFORD, Neil [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). LEE, Soo, Yeun [KR/US]; 40 Westdale Avenue, Daly City, CA 94015 (US). BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). HAFALIA, April, J., A. [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). YAO, Monique, G. [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). RAMKUMAR, Jayalaxmi [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). WALIA, Narinder, K. [US/US]; 890

Davis Street A # 205, San Leandro, CA 94577 (US). LU, Dyung, Aina, M. [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). ARVIZU, Chandra, S. [US/US]; 1706 Morocco Drive, San Jose, CA 95125 (US). ISON, Craig, H. [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US). DING, Li [CN/US]; 3353 Alma Street # 146, Palo Alto, CA 94306 (US). LU, Yan [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). GURURAJAN, Rajagopal [IN/US]; 5591 Dent Avenue, San Jose, CA 95118 (US). WALSH, Roderick, T. [IE/GB]; 8 Boundary Court St. Lawrence Road, Canterbury Kent, Antrim CT1 3EZ (GB). GANDHI, Ameena, R. [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). SWARNAKAR, Anita [CA/US]; 8 Locksley Avenue #5D, San Francisco, CA 94122 (US). FORSYTHE, Ian, J. [US/US]; 308 Roble Avenue, Redwood City, CA 94061 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). AU-YOUNG, Janice, K. [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). ELLIOTT, Vicki, S. [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). LEE, Sally [US/US]; 3643 26th Street, San Francisco, CA 94110 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

## Published:

— with international search report

(88) Date of publication of the international search report:  
11 March 2004

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: KINASES AND PHOSPHATASES

(57) Abstract: The invention provides human kinases and phosphatases (KPP) and polynucleotides which identify and encode KPP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of KPP.

WO 2002/083709 A3

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/10818

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12P 21/06; C12Q 1/68; C12N 15/00, 5/00, 1/20; C07K 1/00; C07H 21/00, 21/04  
US CL : 435/194, 6, 320.1, 252.3, 325; 530/350; 536/23.1, 23.2

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
U.S. : 435/194, 6, 320.1, 252.3, 325; 530/350; 536/23.1, 23.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
MEDLINE EMBASE BIOSISBIOTECHDS SCISEARCH HCAPLUS NTIS LIFESCI BRS/EAST

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages   | Relevant to claim No. |
|------------|--|-----------------------|
| X          | WO 00/21976 A1 (THE GENERAL HOSPITAL CORPORATION) 20 April 2000, (20.04.00), see entire document, especially pages 11-18 and the claims. | 1-7, 9-10, 13, 18     |

☐

Further documents are listed in the continuation of Box C.

☐

See patent family annex.

\* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier application or patent published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "A" document member of the same patent family

Date of the actual completion of the international search

11 February 2003 (11.02.2003)

Date of mailing of the international search report

23 JUN 2003

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231  
Facsimile No. (703)305-3230

Authorized officer  
*Delicia D. Roberts for*  
Maryam Monshipouri

Telephone No. 703 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/10818

## Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claim Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claim Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:  
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-7, 9-10, 12-13, 18, 56-85, directed to SEQ ID NO:1

Remark on Protest

☐  
☐

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet(1)) (July 1998)

# INTERNATIONAL SEARCH REPORT

PCT/US02/10818

## BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

Group I, claim(s) 1-7, 9-10, 12-13, 18, and 56-85, drawn to isolated DNA sequences encoding human kinases and phosphatases, vectors and host cells comprising said products as well as their expression products.

Group II, claim(s) 8, drawn to transgenic organisms comprising said DNA sequences.

Group III, claim(s) 11, 30-43, drawn to antibodies which bind said kinases or phosphatases and methods of use of said antibodies.

Group IV, claim(s) 14-15 and 47, drawn to methods of use of said DNA sequences in a hybridization assay and in a method to generate expression profile of said sequences.

Group V, claims 16-17, drawn to a PCR method using said DNA sequences.

Group VI, claim 19, drawn to methods of treatment using said kinases or phosphatases.

Group VII, claims 20-28, drawn to agonists and antagonists of said polypeptides and methods of treatment using said agonists and antagonists and methods of screening for said agonists and antagonists.

Group VIII, claim 29, methods of assessing toxicity of a test compound using said DNA sequences.

Group IX, claims 44-55, drawn to methods of detecting and purifying polypeptides using antibodies which specifically bind said polypeptides.

Group X, claims 46, 48-55, drawn to arrays comprising said DNA sequences.

For each of inventions of I-X above, restriction to one of the following is also required under PCT Rule 13.1. Therefore election is required of one of the inventions of I-X and one of the inventions of (A)-(O).

- (A). Isolated DNA sequences encoding SEQ ID NO:1.
- (B). Isolated DNA sequences encoding SEQ ID NO:2.
- (C). Isolated DNA sequences encoding SEQ ID NO:3.
- (D). Isolated DNA sequences encoding SEQ ID NO:4.
- (E). Isolated DNA sequences encoding SEQ ID NO:5.
- (F). Isolated DNA sequences encoding SEQ ID NO:6.
- (G). Isolated DNA sequences encoding SEQ ID NO:7.
- (H). Isolated DNA sequences encoding SEQ ID NO:8.
- (I). Isolated DNA sequences encoding SEQ ID NO:9.
- (J). Isolated DNA sequences encoding SEQ ID NO:10.
- (K). Isolated DNA sequences encoding SEQ ID NO:11.
- (L). Isolated DNA sequences encoding SEQ ID NO:12.
- (M). Isolated DNA sequences encoding SEQ ID NO:13.

Form PCT/ISA/210 (second sheet) (July 1998)

# INTERNATIONAL SEARCH REPORT

PCT/US02/10818

(N). Isolated DNA sequences encoding SEQ ID NO:14.

(O). Isolated DNA sequences encoding SEQ ID NO:15. The inventions listed as Groups I-XII and A-O do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical feature of inventions A-O are directed to products of unrelated chemical structure and function. Further, the special technical features of inventions I-III, VI-VII, and X are DNA, transgenic organism, antibodies, polypeptides, modulators and microarrays respectively, which are directed to products of unrelated chemical structure and function. Inventions I, IV, V and VIII share a common special technical feature but are not required to be rejoined under PCT Rule 13.1, because invention I already has a method of use of DNA. Similarly, inventions III and IX share a common technical feature but are not required to be rejoined under PCT Rule 13.1, because invention III already has a method of use of antibodies.

THIS PAGE BLANK (USPTO)